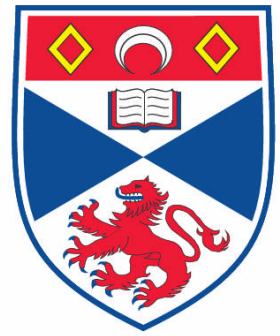
A COMPARATIVE INVESTIGATION OF NUCLEAR DNA CONTENT AND ITS PHENOTYPIC IMPACTS IN *SILENE MARIZII* AND *S. LATIFOLIA*

Mark E. Looseley

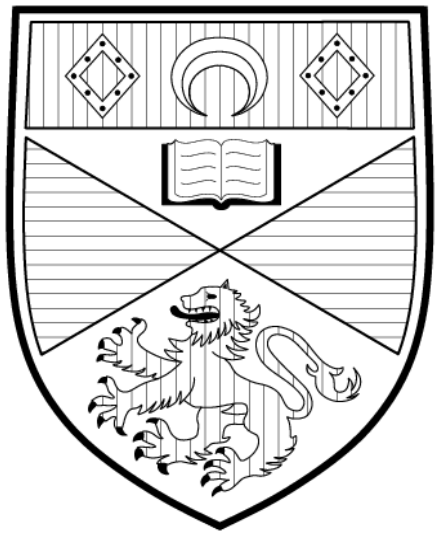
A Thesis Submitted for the Degree of PhD
at the
University of St. Andrews



2008

A Comparative Investigation of Nuclear DNA Content and its Phenotypic Impacts in *Silene marizii* and *S. latifolia*

Mark E. Looseley



Thesis submitted for the degree of Doctor of Philosophy
of the University of St Andrews

25th September 2007

Abstract

Considerable variation exists both within and between species in nuclear DNA content. Despite there being no obvious functional role for much of this DNA, many studies have reported phenotypic correlations with genome size at various taxonomic levels. This suggests that DNA plays a functional role beyond the traditionally understood mechanisms. One such example of a phenotypic correlation with DNA content is present in the genus *Silene*, where a negative correlation between DNA content and flower size exists within and between species. This relationship is consistent with the direction of sexual dimorphism in DNA content (caused by heteromorphic sex-chromosomes) and flower size in the most studied species in the genus: *S. latifolia.*

This thesis takes a comparative approach between two closely related species in the genus (*S. latifolia* and *S. marizii*), which differ markedly in their nuclear DNA content, in order to investigate the nature and phenotypic impacts of variation in DNA content.

A phenotypic survey from a number of *S. marizii* populations reveals that the pattern of DNA content variation in this species is very different to that in *S. latifolia*. In particular, phenotypic correlations with DNA content appear be much weaker, whilst sexual dimorphism in DNA content, when present, appears to occur in either direction. A survey of interspecific hybrids suggests that this may be due to an enlarged *S. marizii* X-chromosome and that DNA content in hybrids may be biased with regard to their parents. Repetitive elements may be significant constituents of plant genomes. A study of Ty1-*copia* class retrotransposons in the two species reveals that they are present as a large and highly heterogeneous population. Phylogenetic analysis of these elements suggests a substantial degree of genetic isolation between the two species. Finally, an assessment of the flow-cytometric method, used to estimate DNA content, reveals substantial error associated with the method, but only limited evidence for stoichiometric effects.

Declaration

I, Mark E. Looseley, hereby certify that this thesis, which is approximately 32 000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.
Date..... Signature of candidate.....

I was admitted as a research student in September 2003 and as a candidate for the degree of PhD in October 2004; the higher study for which this is a record was carried out in the University of St Andrews between 2003 and 2007.
Date..... Signature of candidate.....

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit the thesis in application for that degree.
Date..... Signature of supervisor.....

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration.
Date..... Signature of candidate.....

Acknowledgements

There are a great many people, without whom this thesis would not have been written and to whom I am indebted.

Firstly, thanks to Tom Meagher for suggesting the project, for guidance and supervision during the courses of my research, and for many valuable comments on the draft copies of this thesis. Thanks are due to Harry Hodge for horticultural expertise and for teaching me all I needed to know about flow-cytometry, whilst David Forbes helped me with a wide variety of laboratory techniques. I am grateful to a number of people from within the university for their advice and suggestions during the course of my research. In particular, Daniel Barker for theoretical and practical guidance in the methods of phylogenetic analysis, and Peter Jupp for useful discussions and suggestions regarding statistical techniques.

I was fortunate, also, to receive a great deal of support from outside the university whilst conducting this work. Principally, I would like to thank Andy Flavell and colleagues at S.C.R.I. for their generosity in giving both time and lab space in order to assist me with my work with Ty1-*copia* retrotransposons.

I am grateful to NERC and Beckman-Coulter for funding the research, and providing financial support.

Thanks are also due to my friends and family for their enduring support. In particular, Ineke for endless patience over the years and for doing such a fine job of proof-reading draft versions of this thesis. Also my parents, for supporting me in all that I have chosen to do.

Contents

Chapter 1 Introduction

1.1. Variation in DNA content

1.1.1. Interspecific variation

Even before the discovery of DNA as the molecule responsible for the transmission of hereditary information, it was suggested that genes influenced phenotype by regulating the production of various enzymes (Beadle and Tatum, 1941). The advent of molecular genetics in the late 1950s revealed the mechanism required for this explanation. Indeed, what came to be known as the central dogma of genetics (that DNA, through an RNA intermediate, produces proteins that influence phenotype) still dominates our understanding of the causal relationship between genotype and phenotype. Nevertheless, even as this relationship was being formulated, measurements of DNA content, made in a variety of species, appeared to contradict this neat explanation for the expression of heritable information. The results of these efforts showed that a great deal of DNA content variation exists between species, even when they were closely related, and that this variation appeared to be completely unrelated to an organism's complexity and thus its presumed requirement of genes (Mirsky and Ris, 1951; Cavalier-Smith, 1985). Indeed, the entire 200 000 fold range of genome size found in the eukaryotes is present within the protozoan unicells alone (Biderre et al., 1995; Gregory, 2001). Taken together, these two observations represented a puzzle to early researchers: if DNA was the molecule responsible for the transmission of heritable traits, then why did complex organisms not have more of it than simple ones? This puzzle came to be known as the C-value paradox (C-value being defined as the total DNA

contained in a haploid chromosome set). In more recent times, it has become clear that much of the genome, in many species, is made up of non-functional or duplicated DNA (Flavell et al., 1974; Kubis et al., 1998), and that the size of the genome is not, as was assumed, necessarily related to the number of genes. Indeed, it is probable that more than 90% of the DNA in the biosphere is apparently functionless (Cavalier-Smith and Beaton, 1999), although there is some evidence that some of this DNA may, in fact, be subject to selection (Andolfatto, 2005). Nevertheless, an increased knowledge regarding the nature of genome size variation has removed little of the mystery surrounding the mechanisms by which it arises, how it is maintained, and why some species have larger genomes than others. For these reasons, Gregory (2001) has suggested that the term 'C-value enigma' is more appropriate.

1.1.2. Intraspecific variation

Given that DNA content is a trait that is likely to depend on the cumulative effect of hundreds, perhaps thousands of individual mutations, it is a reasonable expectation that, given sufficient interbreeding, it is also a trait that should remain broadly constant within a species. Nevertheless, there are many reported cases of variation in DNA content below the level of the species (e.g. Teoh and Rees, 1976; Laurie and Bennett, 1985; Vekemans et al., 1996; Rayburn et al., 2004; Smarda and Bures, 2006). Some examples of apparent intraspecific variation in DNA content may be accounted for simply by experimental error (discussed in detail below). Others can be attributed to taxonomic errors, whereby disparate groups are incorrectly classified as a single species. An example of this is in *Scilla bifolia* (Greilhuber, 1979; Greilhuber and Septa, 1985) where a two-fold variation in DNA content within the species vanishes entirely if more stringent taxonomic criteria are applied. Other examples represent what was termed 'orthodox intraspecific variation' by Greilhuber (1998); that is to say, normal chromosomal variation that might be expected to exist within a species, caused, for example, by duplications or deletions; heterochromatic segments; B-chromosomes; polyploidy or even heteromorphic sex-chromosomes in some species. However, even accounting for taxonomic errors and normal chromosomal polymorphism, there are a number of examples of significant variation in DNA content below the species level, even when robust methods for

DNA content estimation have been used (e.g. Reeves et al., 1998; Hall et al., 2000; Moscone et al., 2003).

1.2. Nature of DNA content variation

1.2.1. Causes of DNA content variation

As was noted above, the majority of organisms have a far larger genome than is necessary simply to account for their required complement of genes. In some cases, particularly amongst angiosperms, polyploidy is a common cause of DNA content variation between closely related species. Indeed, it is now thought that all flowering plants are either current polyploids, or have a polyploid history (Kellogg and Bennetzen, 2004). Polyploidy, however, is a far less common phenomenon in animals, and even accounting for polyploidy in plants, there is still considerable variation in DNA content. Most of this residual variation is caused by variation in repetitive DNA (Flavell et al., 1974; Haubold and Wiehe, 2006). Even in the smallest angiosperm genomes, such as *Arabidopsis thaliana* (about 140Mb), more than 20% of the genome is composed of repetitive elements (Arabidopsis-Genome-Initiative, 2002). The term 'repetitive' DNA encompasses a wide variety of sequence types and these can broadly be classified into two major classes. The first of these consists of tandem repeating units. These elements include simple sequence repeats (SSRs), satellite DNA, telomeric repeats and rDNA. The second type of repetitive DNA consists of dispersed repetitive elements, which may take a number of forms; either retroelements or non-retroelement types (Kubis et al., 1998). Retrotransposons, in particular, may be a significant factor in DNA content variation. Little is known about the origin and distant evolutionary history of retrotransposons, but they have subsequently proliferated and may be significant constituents of plant genomes (Kumar and Bennetzen, 1999), representing up to 50% of the nuclear genome in some cases (Bennetzen, 1996; Pearce et al., 1996a; SanMiguel et al., 1996). Furthermore, there is evidence that retrotransposons may rapidly increase in copy number over relatively short evolutionary time-scales. In the *Vicia* genus, for example, Ty1-*copia* copy number per genome varies from 1000 in *V. melanops* to 1,000,000 in *V. faba*. There is also evidence that retrotransposon copy number can vary significantly within a single species (Wendel and Wessler,

2000). Variation in copy number is thought to be a result of variation in rates of element loss rather than variation in levels of retrotransposon activity between species (Navarro-Quezada and Schoen, 2002).

1.2.2. Evolution of DNA content

Large scale accumulation of retroelements (SanMiguel et al., 1996; Bennetzen, 2000), or repeated cycles of polyploidisation (Leitch and Bennett, 1997; Soltis and Soltis, 1999; Otto and Whitton, 2000; Wendel and Wessler, 2000) have both been demonstrated as mechanisms by which genomes may increase in size. Polyploidy in particular is thought to be responsible for the discontinuous variation in genome size that has been observed in a number of taxa (Gregory, 2001). These findings lead to the suggestion that genome size was uni-directional, and that without efficient controls, genomes were destined to increase in size (Bennetzen and Kellogg, 1997). However, it has also been shown that genomes can decrease in size through DNA deletion. Petrov & Hartl (1997) used a phylogenetic approach to examine rates of insertions and deletions in inactive copies of non-LTR retrotransposons in a number *Drosophila* species. They found high rates of DNA loss whilst insertions were almost entirely absent, although this high rate of DNA loss was not consistent across all species, leading to the suggestion that rate of DNA loss may be a significant determinant of genome size (Petrov et al., 2000). Phylogenetic studies of DNA content too, have generally shown that significant decreases in DNA content are just as common as increases (e.g. Caetano-Anolles, 2005; Johnston et al., 2005). Indeed, Wendel (2002), using a phylogenetic approach to infer directionality of genome size changes in *Gossypium* species, concluded that genome size decreases were actually a more common occurrence in the history of these species than increases. Although the mechanisms for significant DNA loss are less well understood than those for the large scale accumulation of DNA, a number of studies have revealed the diversity and extent of mechanisms by which genomic deletions may occur (Petrov et al., 1996; Petrov and Hartl, 1997; Vicient et al., 1999; Kirik et al., 2000; Shirasu et al., 2000).

Nevertheless, despite the increased understanding of how genomes change in size, the reasons why are less well known. In order to understand the evolutionary forces

that are driving genome size evolution, it is important to study the consequences of DNA content variation.

1.3. Phenotypic correlations with DNA content

1.3.1. Correlations

Despite the widely held view that 'non-functional' DNA is of no phenotypic significance, correlations have been observed between DNA content and a number of cellular and organismal traits. An almost ubiquitous and particularly well studied example of this is the correlation between DNA content and cell size that has been observed in almost all taxa examined (e.g. Cavalier-Smith, 1978, 1982; Hardie and Hebert, 2003). As stated in Gregory (2001), strong correlations are also common between DNA content and both mitotic and meiotic rates of cell division (e.g. Van't Hof and Sparrow, 1963; Van't Hof, 1965; Bennett, 1971; Evans et al., 1973; Nagl, 1974; Grosset and Odartchenko, 1975; Price and Bachmann, 1976).

In addition to these cellular traits, correlations between DNA content and whole organism traits have also been reported. These include body size (Gregory et al., 2000), seed size (Chung et al., 1998), vegetative characters (Caceres et al., 1998), metabolic rate (Vinogradov, 1995, 1997; Olmo, 2003), flower size (Meagher and Costich, 1996; Meagher and Vassiliadis, 2005) and life history traits (Bennett, 1972; Grime and Mowforth, 1982), e.g. the relationship between generation time and DNA content in herbaceous plants (Bennett, 1972).

Environmental and ecological correlations with DNA content, both between and within species have also been reported (Garnatje et al., 2004; Knight et al., 2005; Murray, 2005) in particular with respect to altitude (Rayburn and Auger, 1990; Reeves et al., 1998; Temsch and Greilhuber, 2000) and latitude (Bennett, 1976; Levin and Funderburg, 1979; Ohri et al., 1998; Temsch and Greilhuber, 2001).

1.3.2. Proposed explanations for phenotypic correlations

Attempts to explain the correlations described above (particularly the correlation between DNA content and cell volume) can broadly be divided into those that explain DNA content as coincidental to phenotype and those which see phenotype as being dependent on DNA content.

The first class of theories are known as mutation pressure theories (Gregory, 2001). Under these theories, the genome is assumed to be under constant pressure to increase in size. The mechanism of increase may be neutral, whereby non-coding DNA increases by chance, termed 'junk DNA'(Pagel and Johnstone, 1992), or due to intragenomic selection on self-replicating genomic elements, termed 'selfish DNA' (Doolittle and Sapienza, 1980; Orgel and Crick, 1980). Under both these theories selection against DNA content occurs when the tolerance of the cell for DNA content is exceeded. Thus, a cell's capacity for DNA is seen as dependent on its volume and the correlation with DNA content follows from this.

The second class of theories are known as the optimal DNA theories (Gregory, 2001). These envisage DNA content as being directly related to phenotype. This effect may be co-evolutionary, as is the case under the 'nucleoskeletal' theory (Cavalier-Smith, 1978), or causative, as under the 'nucleotype' theory, proposed by Commoner (1964) and developed by Bennett (1971) and Bennett (1972). Under the nucleotype theory, DNA content has a direct causative effect on cellular parameters that is entirely independent of the information content of the DNA. As such, DNA content is regulated directly by the action of selection on cellular or organismal characters.

All of the theories described above directly address the issue of the correlation between DNA content and cellular traits. However, under each of these theories (with the possible exception of the nucleotype theory), the relationship between DNA content and other traits are less clear. Cellular parameters may have important consequences for a wide range of whole organism traits, but it is not obvious that this is sufficient to explain the many traits that have been reported to correlate with DNA content.

A more recently proposed theory, under which DNA content may independently influence a wide variety of traits, is the genome regulation theory (Meagher and Vassiliadis, 2005). Under this theory, the accumulation of repetitive DNA has localised or general effects on gene regulation. One such mechanism proposed by Vinogradov (1998) sees excess DNA playing a 'buffering' role through the non-specific binding of proteins, affecting overall metabolic rates. Dispersed repetitive elements may affect phenotype more directly, with retrotransposons regulating the expression of specific genes by acting as *cis* regulators (White et al., 1994) or more generally, by causing the build-up of *trans* acting expression factors in genomic regions where they occur (e.g. Kato et al., 1999).

1.4. Effects of DNA content on reproductive isolation

Under the genome regulation theory, much of the variation in DNA content may reflect the action of selection on specific genetic elements through their effects on phenotype rather than DNA content as a whole. Thus, it would be expected that differences between species in DNA content (and the phenotypic characters these affect) may be attributable to variation in specific genetic elements. This variability might be expected to contribute to interspecific differences in recombination rates (e.g. Williams et al., 1995), which in turn might be expected to influence interspecies incompatibility and hybrid breakdown. An additional theory as to the functional significance of DNA content variation is that it may help to determine homologous chromosome compatibility during meiosis (Irick, 1994). Thus, it might be expected that variation in DNA content could contribute towards mismatch. However, few studies have directly examined the effect of genome size variation on reproductive isolation either within or between species. Those studies that have, have generally found that relatively large differences in DNA content have little effect on reproductive isolation (e.g. Rees and Narayan, 1981; Galloway and Etterson, 2005). Even so, such assessments are difficult in practice as large scale difference in DNA content may equally well be an effect, rather than a cause, of reproductive isolation (Murray, 2005).

1.5. Measuring DNA content

Apart from in the very few species for which complete genome sequence is available, most studies of DNA content must rely on indirect methods of estimation. Traditionally, this has been done either by the process of Fuelgen mirospectrophotometry or by flow-cytometry with an appropriate fluorochrome. Both methods rely on the comparison of stained sample nuclei with those of an appropriate internal standard, which is assumed to account for variation in dye intensity, or staining inhibitors. Nevertheless, a growing body of work suggests that the action of cytosolic staining inhibitors may have differential effects on sample and standard nuclei (e.g. Price et al., 2000; Noirot et al., 2003; Greilhuber, 2005). Thus, observed variation in estimates of DNA content may simply reflect stoichiometric error rather than an underlying variation in genome size. For this reason, all such methods of estimating DNA content should be subject to a critical assessment, before true variation in DNA content is concluded.

1.6. Study species

1.6.1. *S. marizii* and *S. latifolia*

The work described in this thesis focuses on two species in the genus *Silene* (Caryophyllaceae): *S. marizii* and *S. latifolia*. Both species are members of the section *Elisanthe*, and both have the same chromosome number (2n=24), hybridising readily under glasshouse conditions despite maintaining separate identities in wild populations. This suggests that species boundaries in the two plants may be maintained by ecological separation with strong selection against hybrids (as has been reported in the similar *S. latifolia*/ *S. dioica* system (Goulson and Jerrim, 1997)). The two species vary dramatically in their distributions; *S. latifolia* is native to Europe, where it shows a widespread distribution, and has been naturalised in North America. *S. marizii*, on the other hand is an Iberian endemic, restricted to northern Portugal and adjacent regions of Spain (Talavera, 1990). Both species are dioecious (Talavera, 1990) and *S. latifolia* in particular is well known in the scientific literature for being one of the few dioecious plant species with heteromorphic sex-chromosomes (XX females, XY males). Sex-chromosome differentiation in *S. latifolia* has been particularly well studied (Westergaard, 1958;

Ciupercescu et al., 1990; Charlesworth, 2002; Grabowska-Joachimiak and Joachimiak, 2002; Lengerova et al., 2003), with heteromorphy being so pronounced that differences in DNA content are detectable between males and females (Costich et al., 1991; Dolezel and Gohde, 1995), with males showing elevated DNA content due to an enlarged Y-chromosome relative to the X-chromosome.

1.6.2. Sex-chromosome evolution in *Silene*

Sex-chromosomes in *Silene* are thought to have arisen recently (between 10 and 20 million years ago) (Filatov and Charlesworth, 2002) and are, therefore, particularly valuable for the study of early processes of sex-chromosome evolution. In *S. latifolia* it is thought that heteromorphy between X and Y-chromosomes developed following gradual suppression of recombination between the two chromosomes. A number of population genetic processes are thought to result in the accumulation of repetitive DNA (both tandem and dispersed) in non-recombining regions of the Y-chromosome (Charlesworth et al., 1994), Indeed, the enlarged Y-chromosome, in *S. latifolia*, has been attributed to a build up of repetitive DNA (Grant et al., 1994) and chloroplast sequences (Kejnovsky et al., 2006) in such regions, although there is no evidence for an accumulation of LTR retrotransposons within the sex-chromosomes (Matsunaga et al., 2002).

1.6.3. Floral morphology

In both species, the sepals subtending the petals are fused to form a calyx tube. The diameter of the calyx tube along with limb length and claw length have previously been used as measures of flower size in studies of *S. latifolia*. These studies have shown that in addition to the sexual dimorphism in DNA content, sexual dimorphism in flower size is also present within the species (Meagher, 1992, 1994; Meagher and Costich, 1994), with females showing larger flowers than males (Figure 1.1).

Flower size is a character of clear ecological significance that may have important consequences on floral attractiveness and, therefore, rates of pollinator visitation (Bell, 1985; Delph et al., 1996). Within dioecious angiosperms there is the potential

for sex-specific selection on flower size and the resulting flower size dimorphism has often been interpreted in terms of Bateman's principle (Bateman, 1948), which suggests that, due to the greater variance in reproductive success within males compared to females, males will be under stronger selection for floral attractiveness than females. In European populations of *S. latifolia,* pollination biology is complicated by their association with moths of the genus *Hadena* (Leptidoptera: Noctuidae), which act both as seed predators (Brantjes, 1976) and pollinators (Jurgens et al., 1996). Calyx diameter (one measure of flower size) has previously been identified as a factor in protection from seed predators in a related *Silene* species (*S. vulgaris*) (Pettersson, 1991). A study of natural populations of *S. latifolia* has identified some evidence for sex-specific selection within the species, but significant heterogeneity in patterns of selection between years and populations have prevented any firm conclusions about the origin of sexual dimorphism in flower size within the species (Wright, 2000; Wright and Meagher, 2004).

1.6.4. Correlation between DNA content and flower size

Further studies of phenotypic variation in *S. latifolia* have shown a significant negative correlation between DNA content and flower size. This relationship is consistent with the direction of sexual dimorphism in the species, but is also present within the sexes (Meagher and Costich, 1994). Further work revealed that this relationship was strongest when an AT-biased measure of DNA content was used (Meagher and Costich, 1996), suggesting that repetitive DNA, specifically, may be responsible for the relationship; plant repetitive DNA often being composed of AT-rich sequences (Lagercrantz et al., 1993; Wang et al., 1994). In addition to the studies of phenotypic correlations, it has been demonstrated that these are underpinned by strong genetic correlations, with DNA content showing a correlated response to selection on flower size (Meagher and Costich, 1996) (Figure 1.2). Comparisons of cell sizes between males and females of *S. latifolia* have shown that the sexual dimorphism in flower size is a consequence of greater numbers of cell divisions rather than greater cell sizes. This is consistent with previous studies which have shown that smaller genomes are associated with faster rates of mitotic cell division (Gregory, 2001).

=

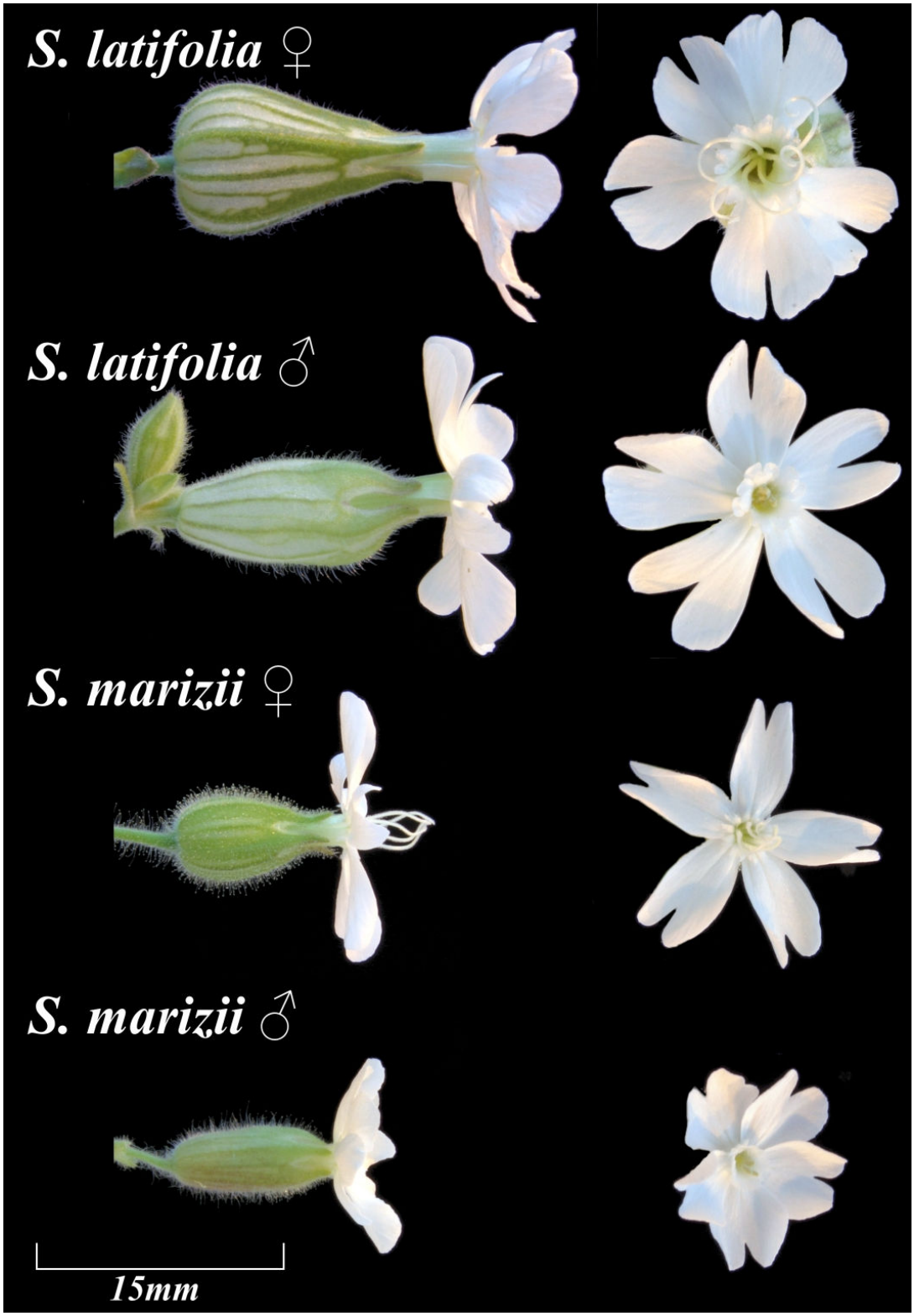


Figure 1.1: Male and female flowers from *S. latifolia* and *S. marizii*, illustrating both the sexual dimorphism and the interspecific variation in flower size (photography by Sean Earnshaw: School of Biology, University of St Andrews).

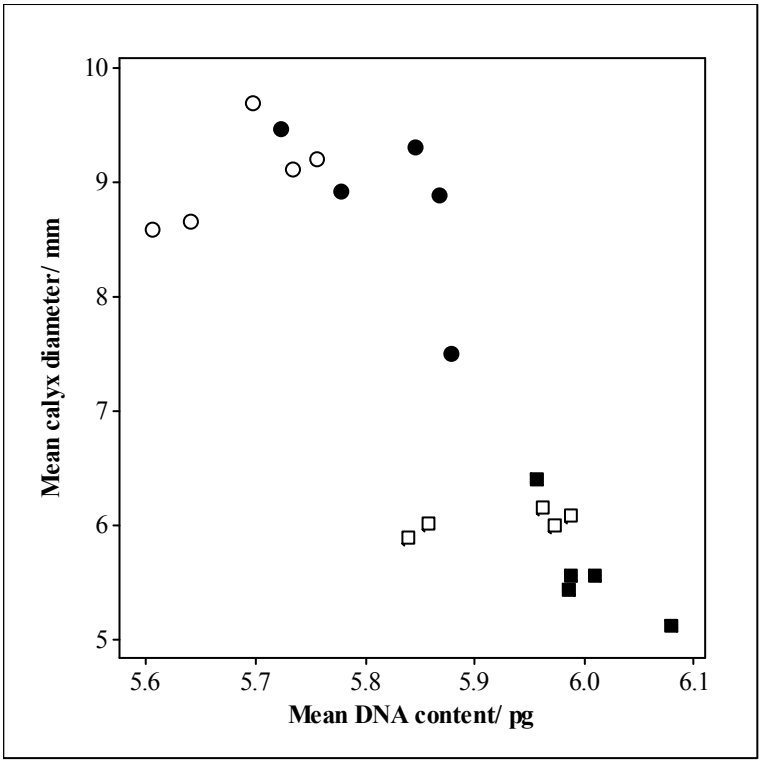


Figure 1.2: Relationship between mean calyx diameter in and DNA content amongst *S. latifolia* lines selected for either large or small calyces, illustrating the correlated response in DNA content (Meagher and Costich, 1996). Squares represent males and circles females. Open and filled circles indicate selection lines taken from either of two natural populations.

1.6.5. Correlations within the section *Elisanthe*

In addition to the work examining the relationship between DNA content and flower size in *S. latifolia*, the generality of this relationship has been tested in four species of the section *Elisanthe* (*S. latifolia*, *S. marizii*, *S. diclinis* and *S. dioica*) (Meagher and Costich, 2004). This study found that the sexual dimorphism in DNA content was indeed consistent across all the species, and in addition, detected significant variation in DNA content between the species, and between populations of the same species. There was also a trend towards a negative correlation between DNA content within sexes of each species as well as within species as a whole.

1.6.6. *S. marizii* and *S. latifolia* as a model system

S. marizii has substantially smaller flowers than *S. latifolia* and a significantly greater DNA content (Meagher and Costich, 2004). These differences, along with the ease of hybrid formation, mean that *S marizii* and *S latifolia,* along with interspecific hybrids between the two species, may provide a useful model system in which to study the genomic distribution of interspecific differences in genome size,

and the mechanism by which variation in DNA content influences phenotype. Thus, the overall objective of this thesis is to utilise this model system to explore the genomic dynamics of variation in nuclear DNA content and its impact on phenotypic variation.

1.7. Thesis outline

In this general introduction, an outline has been given of the extent of DNA content variation, both within and between species, along with a summary of the accepted mechanisms by which such variation might arise. Associated phenotypic correlations with DNA content variation have also been introduced. The specific example of the negative correlation between DNA content and flower size in the *Silene* genus has been described and a potential model system, through which to investigate the nature and cause of this relationship (hybrids between *S. latifolia* and *S. marizii*), has been proposed.

The second chapter describes work that investigates the extent of variation in DNA content and flower size within a number of Portuguese populations of *S. marizii*. Phenotypic correlations between these characters are examined, along with patterns of sexual dimorphism. These results are compared with patterns previously reported in populations of *S. latifolia*.

The third chapter examines the process of hybridisation between *S. marizii* and *S. latifolia*. In particular, rates of germination in a number of populations of hybrid seed are examined as is the extent to which variability in germination rates might bias the traits of interest in these populations. Patterns of DNA content variation in each generation of an F2 crossing design are examined and correlations between DNA content and flower size examined in the F2.

The potential role of specific repetitive elements in influencing DNA content and their influence over associated phenotypic correlates are of considerable interest. The fourth chapter examines a new method for estimating genomic copy numbers of Ty1-*copia* class retrotransposons in large numbers of individuals. The site-specific

amplification polymorphism (SSAP) method examined here is compared with another PCR based method that has recently been developed.

Repetitive DNA plays a significant role in DNA content variation both within and between species. As such, the evolutionary dynamics of repetitive elements are of great interest. In the fifth chapter, a phylogenetic analysis of Ty1-*copia* RNaseH sequences from *Silene* species and a number of other species is made. The results are examined with regard to previous phylogenetic analyses in other Ty1-*copia* genes. Patterns of heterogeneity within conspecific sequences are compared to overall DNA content and estimates of Ty1-*copia* copy number with regard to theoretical predictions. The results of the phylogenetic analysis are also used to examine the likely degree of genetic isolation between *S. marizii* and *S. latifolia*.

Crucial to all the work examining patterns of DNA content variation is an accurate and unbiased method for estimating DNA content. Work described in the sixth chapter aims to investigate the potential sources of error and bias in the flow-cytometry protocol used to estimate DNA content in much of the work described in this thesis. Random error associated with the measurement of the internal standard as well as stoichiometric error is considered and the likely extent of such effects in DNA content data sets quantified.

Chapter 2 *Silene marizii* phenotypic survey

2.1. Introduction

2.1.1. Evolution of flower size

Flower size is a quantitative trait that is an important component of floral displays in animal pollinated plant species and may, therefore, have important consequences for reproductive success (Bell, 1985). An interesting feature of dioecious angiosperms is the potential for sex-specific selection on floral traits (in particular flower size) (e.g. Meagher, 1994; Wright and Meagher, 2004). This in turn may cause a significant sexual dimorphism in these traits. Such sexual dimorphisms in flower size are widespread (Delph et al., 1996) and have often been interpreted in terms of Bateman's principle (Bateman, 1948). Bateman's principle, when applied to entomophilous plants (Bell, 1985), suggests that the disparity between male and female gametes (pollen grains far outnumbering ovules) will cause male reproductive success to be more dependent on floral attractiveness, and thus rates of insect visitation, than female reproductive success, which may be more resource dependent. Furthermore, in many species excessive floral displays may actually reduce female fitness through the attraction of seed predators, which, in some species, may also be pollinators. Indeed, this is the case in *S. latifolia*, where moths of the genus *Hadena* may act as pollinators (Jurgens et al., 1996) and also as seed predators (Brantjes, 1976). A further functional role played by flowers is to protect developing reproductive structures and ovules (in female flowers). If the protective

role is a significant driver of flower size, then sexual dimorphism may act in either direction, depending on which sex contains the larger reproductive structures. Studies of dioecious angiosperms suggest that it is a balance between these two functional roles which determines patterns of sexual dimorphism in flower size (Meagher, 1992; Delph et al., 1996; Meagher, 1999).

2.1.2. Sexual dimorphism in *S. latifolia*

A well studied example of sexual dimorphism in flower size exists within *S. latifolia*. An interesting feature of this example is an associated negative correlation between flower size and nuclear DNA content. Intraspecific variation and sexual dimorphism in both DNA content and floral characters has been well documented in *S. latifolia* (Costich et al., 1991; Meagher and Costich, 1994) with males tending to have smaller flowers and larger genomes than females. This sexual dimorphism in DNA content has been attributed to the presence of heteromorphic sex-chromosomes in the species (XX female; XY male), with greatly enlarged Y-chromosomes relative to the X-chromosome (Westergaard, 1958; Ciupercescu et al., 1990; Costich et al., 1991; Dolezel and Gohde, 1995). The negative correlation between floral characters (particularly calyx diameter) and DNA content has also been well studied in the species (Meagher and Costich, 1994) and holds within, as well as between, the sexes. Additional work has shown that selection on flower size produces a correlated response in DNA content (Meagher and Costich, 1996) and that phenotypic correlations both within and between species are a general trend across *Silene* section *Elisanthe* (Meagher and Costich, 2004).

This pattern of sexual dimorphism in flower size in *Silene* appears contrary to Bateman's principle, which would predict larger male flowers. However, males tend to produce more flowers (Meagher, 1992, 1994; Steven et al., 2007), and, therefore, male floral displays may be larger despite the smaller size of each individual flower. In addition to this, one of the three floral characters (calyx diameter) has been shown to serve a protective rather than attractive role (Pettersson, 1991). Correlated sexual dimorphisms in other floral traits (limb length and claw length) have been attributed to strong genetic correlations between these

and calyx diameter (Meagher, 1992, 1999), suggesting that the potential for independent sex-specific evolution of these three floral traits is limited.

Traditional quantitative genetic methods aim to partition phenotypic variance into genetic and environmental components, with genetic components of variance being considered at a genome wide level. Advances in genomic techniques have allowed direct examination of the underlying molecular basis for this variation to be made. Another species in the *Silene* genus, *S. marizii*, is a close relative of *S. latifolia* and the two species hybridise readily, yet are very different in terms of flower size and nuclear DNA content. As such, this species pair may provide an ideal model system with which to study the genomic basis of DNA content/ flower size variation within the genus.

2.1.3. Aims

Crucial to the interpretation of any study involving interspecific hybrids is a thorough understanding of patterns of phenotypic variation within the parental species. Despite the large body of work examining intraspecific variation in DNA content and floral characters in *S. latifolia*, the same cannot be said of *S. marizii*, which is a much less common species with fewer populations and a much more restricted distribution. As such, prior to any genomic study using hybrids of the two species it is important to conduct a more comprehensive survey of phenotypic variation in both DNA content and flower size. The aim of the current study was to conduct a glasshouse survey of variation within populations of *S. marizii* sampled in northern Portugal and to compare patterns of variation with those previously identified in populations of *S. latifolia*. The study examined rates of germination amongst natural populations of *S. marizii* and sex-ratios amongst these plants. In addition, the extent and pattern of variation in DNA content and flower size were studied, and relationships between these traits examined. The results were compared to patterns previously reported in *S. latifolia*.

2.2. Methods

2.2.1. Identification of populations

Seed from five populations of *S. marizii* were collected during a week long field trip to Northern Portugal in July 2004. Seed collections were made from one previously identified population near the town of Mangualde (Table 2.1; Figure 2.1). Additional populations were identified from herbarium records in the herbarium at the University of Porto. From these records, *S. marizii* plants were identified in a further four locations (Table 2.1; Figure 2.1). From each of the populations up to ten maternal plants (or as many as were available) were located, and seed collected (Table 2.1). In all locations, the density of *S. marizii* individuals was low.

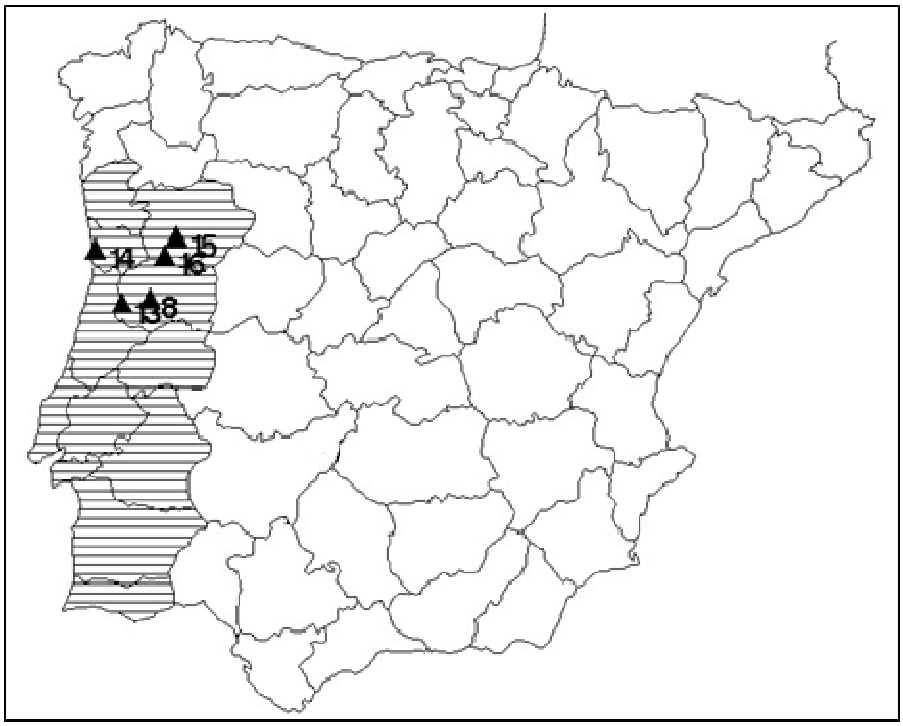


Figure 2.1: Map of Portugal (shaded) and Spain showing the location of the *S. marizii* populations. Population numbers correspond to those given in Table 2.1.

2.2.2. Cultivation of plant material

Collected seeds were transported to the University of St Andrews, where they were germinated and cultivated under glasshouse conditions. Ten seeds per maternal plant were sown in 3 cm root trainers. Seeds were planted in a fully randomised design arranged in groups of 50 seed (5x10) per root trainer tray. Seeds were kept

moist by automatic sprinklers, and germinated under 400W mercury vapour lamps with a photo period of 16 hours. When germinated seedlings were of sufficient size, they were transplanted to 11.5cm pots and, again, arranged in a fully randomised array. Plants were left to grow until the first flower reached floral maturity. At this point morphometric measurements were made on up to four flowers. Tissue for flow-cytometric analysis was not collected until morphometric measurements were completed. Due to low germination rates in many of the populations, a second block of seeds were sown six months after the first. Again, 10 seeds per maternal plants were sewn and raised under identical conditions to those described above.

Table 2.1: Location and details for each of the *S. marizii* population identified in the study. The collection date and the number of maternal plants from which seed were sampled are also shown.

Name	Location	Altitude/ m	Population Number	Collection Date	Maternal Plants
Mangualde	40.616N 7.740W	570	8	11/07/2004	10
Porto	41.141N 8.601W	10	14	13/07/2004	10
Tabuaço	41.096N 7.557W	788	16	14/07/2004	10
El Caramulo	40.548N 8.201W	1056	13	12/07/2004	8
San Lourenço	41.293N 7.374W	214	15	14/07/2004	10

2.2.3. Morphometric measurements

Three floral characters were measured (Figure 2.2), namely: calyx diameter, limb length and claw length. Measurements were made with a set of vernier callipers to the nearest 0.1mm. Limb length was measured for one randomly selected petal on each flower and the measurement for calyx diameter was made across the widest part of the calyx.

2.2.4. Flow-cytometry

A flow-cytometric approach was used to estimate DNA content. Young leaves were collected from the plants that were to be assayed and rinsed in distilled water to remove any contaminants. Of the cleaned leaf tissue, 50mg was weighed out and placed in a plastic petri dish (35mm x 10mm), which was placed on ice. Extraction buffer (1ml) (10mM MgSO4, 50mM KCl, 0.247% Triton X-100, 0.099% dithiothreitol) containing propidium iodide (0.01%) was added to the petri dish. The leaf tissue was finely sliced with a sharp scalpel and left for 2 minutes to allow

the nuclei to come into suspension. The suspension containing the nuclei was filtered through a 30µm mesh (to remove any residual leaf material) into a 1.5ml micro-centrifuge tube. The suspension was centrifuged for 30 seconds at 13 000 rpm to collect the stained nuclei. The supernatant was discarded, and the nuclei re-suspended in 300µl extraction buffer containing propidium iodide, chicken red blood cells (which were used as internal standards) and 0.4ml RNAse. The suspension was incubated at 37°C for 15 min.

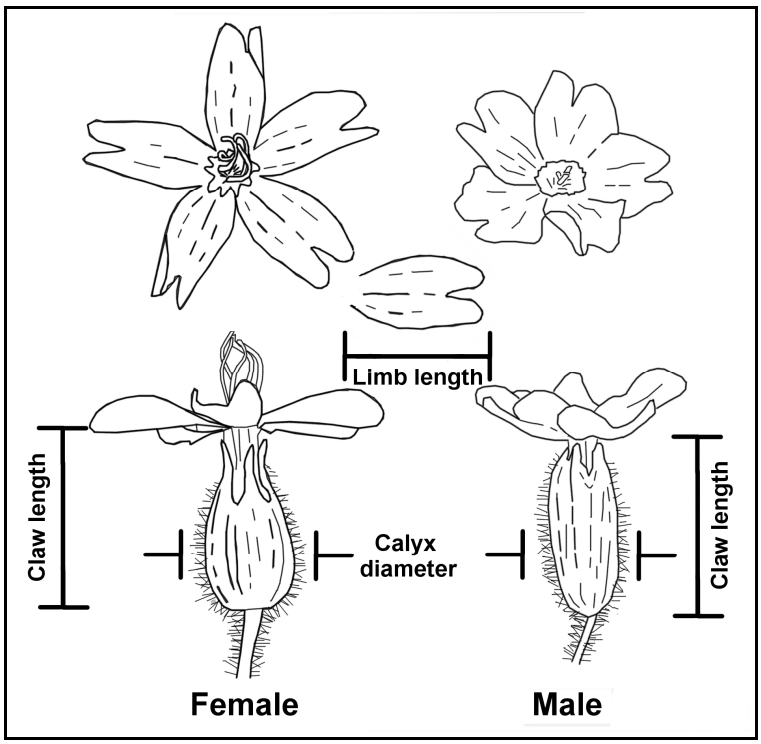


Figure 2.2: Drawing of a male and female *S. marizii* flower, showing the three floral traits measured: claw length, limb length and calyx diameter.

Flow-cytometric assays of the stained nuclei were conducted using a Coulter Epics XL-MCL flow cytometer (Beckman Coulter); the analysis was conducted using the Beckman Coulter EXPO32 software. Nuclei were passed in front of a 15mW Argon ion laser emitting at 488nm. The forward scattering of the beam (scattering at narrow angles to the direction of the laser) was measured by a photo voltaic sensor. In addition, the fluorescence of the propidium iodide was determined by passing the light generated at a wide angle to the beam through a 600nm dichroic long pass filter and a 620nm band-pass filter (transmits light between 605nm and 635nm) to another sensor. Samples were run for either 120s, or until 10 000 nuclei (including chicken red blood internal standards) had been analysed. Samples for

which fewer than 8 000 nuclei had been analysed after 120 seconds were reanalysed. Doublet nuclei (particles composed of several amalgamated nuclei, and thus showing inflated fluorescence relative to single nuclei) were discarded by selecting only the nuclei in which the peak fluorescence was high compared to an integrated measure (total florescence). The remaining nuclei consisted of chicken red blood nuclei, diploid nuclei, and tetraploid nuclei (from cells undergoing mitotic division). The mean and mode fluorescence of the chicken red blood standard and diploid nuclei groups were measured, along with the number of nuclei in each group and the coefficient of variation. DNA content was estimated as $g = 2.33\left[\frac{s}{c}\right]$, where g is the estimated DNA content in pg, s is the mean fluorescence of the sample nuclei and c is the mean fluorescence of the chicken standard. 2.33 is the known nuclear DNA content of chicken red blood cells (in pg) (Arumuganathan and Earle, 1991). This method is illustrated in Figure 2.3. An analysis of error associated with this method suggests that, whilst error in estimates of DNA content may be large, these estimates are unbiased. This analysis is described in Chapter 6.

2.2.5. Statistical analysis

All statistical analyses were conducted using Minitab 15 software (Minitab-Ltd, 2006). Sex-ratios amongst flowering individuals were compared to the expectation of 1:1 (males: females) using a chi-squared goodness of fit test. Sex-ratios were tested overall as well as for each population. Similarly, germination rates between maternal sibships within each population were tested using a chi-squared goodness of fit test against the expectation of equal germination rates.

For analyses using floral traits, mean values across all flowers within an individual were used. Factors affecting patterns of variation in the three floral traits and in DNA content variation were studied using Analysis of Variance (ANOVA). The effect of population, sex and a population by sex interaction were examined. For factors showing a significant effect, a Tukey multiple comparison test was used to examine differences in the mean between each population.

Correlations between variables were examined using Pearson correlation coefficients. Overall correlation coefficients were calculated along with correlations within sexes and populations.

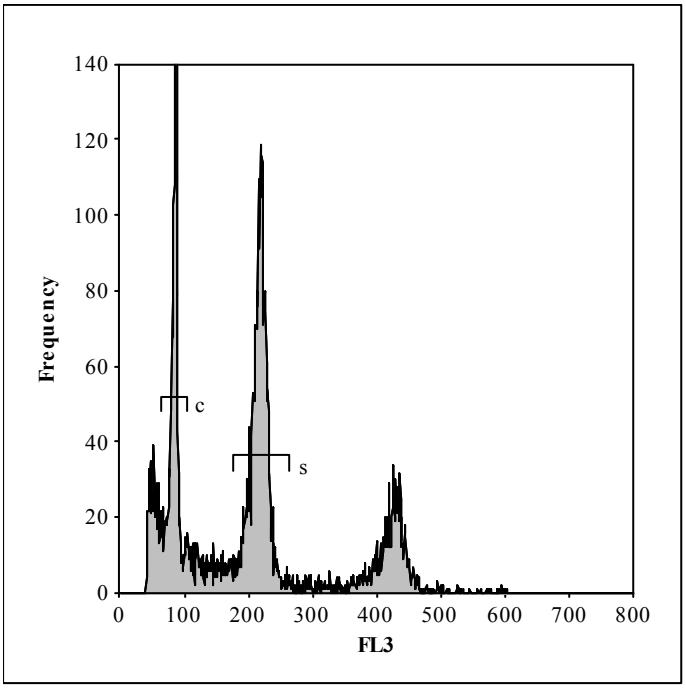


Figure 2.3: Example frequency histogram from a flow-cytometry run. The y-axis shows the number of nuclei falling within each integer value of fluorescence (FL3, x-axis). Horizontal bars indicate the approximate position of standard nuclei (c) and the sample nuclei (s).

2.3. Results

2.3.1. Germination rates

Germination rates were low in most populations, ranging from only 10% in population15, to 26% in population 8 (Table 2.3). This difference in germination rates amongst populations was highly significant (χ^2=20; p<0.001). Within most populations there was also evidence for unequal germination rates between seed samples from different maternal plants (Table 2.2). Flowering rates amongst germinated seed were also low (around 50% in most populations) (Table 2.3).

Table 2.2: Chi-squared goodness of fit test for equality of germination rates between maternal sibships within each population.

Population	χ^2	*p*
13	20	0.02
8	27	0.00
14	14	0.09
15	15	0.04
16	7	0.64

Table 2.3: Germination rates and sex-ratios in the survey population. Chi-squared values and *p*-values show the goodness of fit between observed sex-ratios and expectations of equality.

Pop	Seed Sown	Germinated	Germination rate	Flowered	Sex-ratio (males/total)	χ^2	p
8	200	51	0.26	29	0.66	2.79	0.095
14	200	47	0.24	23	0.65	2.13	0.144
16	200	16	0.08	8	0.50	0.00	1.000
13	200	21	0.11	10	0.80	3.60	0.058
15	200	20	0.10	9	0.67	1.00	0.317
Overall	1000	155	0.19	79	0.66	7.91	0.005

2.3.2. Sex-ratios

A notable feature of this study is the excess of males compared to females in the survey population, a disparity which was highly significant (Table 2.3). This inequality was present in all but one of the populations examined, although within populations, there was only weak evidence for an excess of males.

2.3.3. DNA content variation

There was no evidence to suggest an overall sexual dimorphism in DNA content when individuals were pooled across all populations. However, there was evidence for a difference between populations (Table 2.4). When comparisons between populations were made, there was evidence for a difference only between the means of population 8 and population 14 (Table 2.5). There was also strong evidence for a population by sex interaction, suggesting that sexual differences in DNA content varied between populations. Indeed, when the within population sex means are compared, there is strong sexual dimorphism within two of the populations. In population 8, males showed a higher DNA content, whilst in population 14 females had larger genomes. Sexual dimorphism in each of the other populations was much lower (Figure 2.4).

2.3.4. Floral variation

There was strong evidence for population level differences for all three of the floral traits examined (Table 2.4), although these differences were only significant in comparisons between a few populations (Table 2.5). There was also evidence for a sexual dimorphism in calyx diameter and limb length, and weak evidence for sexual dimorphism in claw length. In no case was there a significant interaction between population and sex, with the direction of sexual dimorphism for each character appearing to be generally consistent across populations (although several populations do not appear to be sexually dimorphic). Males tend to have larger limbs and claws and smaller calyces than females (Figure 2.4 a,b,c), although calyx diameter was the only trait to show strong and consistent evidence for a sexual dimorphism.

2.3.5. Correlations

When overall correlations (pooled over sexes and populations) between the traits were considered, the only highly significant correlation is a positive relationship between limb length and claw length (Table 2.6). One other correlation (DNA content and claw length) showed a trend towards a negative relationship.

When the sexes are considered separately, it becomes clear that the positive correlation between limb and claw length is much stronger in females than in males. Similarly, the negative correlation between DNA content and claw length appears somewhat more strongly in females than in males. A number of other correlations appear significant when sexes are considered individually that do not appear in the pooled analysis. The most notable of these is a significant positive correlation between claw length and calyx diameter in males.

Correlation analyses conducted separately on populations or sexes within populations should be treated with caution due to the sample sizes in some of the populations and the generally low numbers of female plants. However it can be noted that in many cases (particularly the relationship between DNA content and limb length) significant correlations appear only in certain populations (Table 2.6).

Table 2.4 ANOVA tables for the three floral traits and nuclear DNA content examining the effects of population and sex.

Source	Degrees of Freedom	Sum of Squares	Adjusted Sum of Squares	Adjusted Mean Square	*F*-ratio	*p*
a. DNA Content (pg)						
Population	4	9.73	5.66	1.41	2.8	0.033
Sex	1	0.00	0.02	0.02	0.03	0.86
Population x Sex	4	6.27	6.27	1.57	3.11	0.021
Error	65	32.78	32.78	0.50		
Total	74	48.78				
b. Limb Length (mm)						
Population	4	22.79	16.69	4.17	2.59	0.044
Sex	1	15.13	6.88	6.88	4.28	0.043
Population x Sex	4	6.30	6.30	1.57	0.98	0.426
Error	65	104.58	104.58	1.61		
Total	74	148.80				
c. Claw Length (mm)						
Population	4	35.06	27.41	6.85	3.95	0.006
Sex	1	15.04	6.60	6.60	3.81	0.055
Population x Sex	4	9.59	9.59	2.40	1.38	0.250
Error	65	112.74	112.74	1.73		
Total	74	172.44				
d. Calyx Diameter (mm)						
Population	4	10.69	13.96	3.49	4.86	0.002
Sex	1	31.31	32.14	32.14	44.74	<0.001
Population x Sex	4	3.80	3.80	0.95	1.32	0.271
Error	65	46.68	46.68	0.72		
Total	74	92.48				

Table 2.5: Tukey's test for population level differences between each population for DNA content and the three flower size measurements. Upper values for each comparison show the difference in the mean and lower values are adjusted *p*-values.

a. DNA content (pg)

	8	**14**	**15**	**16**
13	0.498	-0.123	-0.183	0.105
	0.523	0.996	0.989	0.999
8		-0.62	-0.68	-0.393
		0.035	0.138	0.699
14			-0.06	0.227
			0.999	0.949
15				0.288
				0.936

b. Limb length (mm)

	8	**14**	**15**	**16**
13	-0.217	0.009	1.45	0.052
	0.996	>0.999	0.231	>0.999
8		0.226	1.667	0.269
		0.976	0.018	0.986
14			1.441	0.044
			0.07	>0.999
15				-1.398
				0.211

c. Claw length (mm)

	8	**14**	**15**	**16**
13	-2.133	-1.673	-0.926	-1.309
	0.003	0.042	0.66	0.321
8		0.461	1.208	0.824
		0.742	0.152	0.515
14			0.747	0.363
			0.635	0.961
15				-0.383
				0.976

d. Calyx diameter (mm)

	8	**14**	**15**	**16**
13	-1.327	-1.746	-1.563	-1.035
	0.009	0.001	0.011	0.18
8		-0.42	-0.236	0.291
		0.481	0.962	0.922
14			0.1831	0.711
			0.987	0.299
15				0.528
				0.748

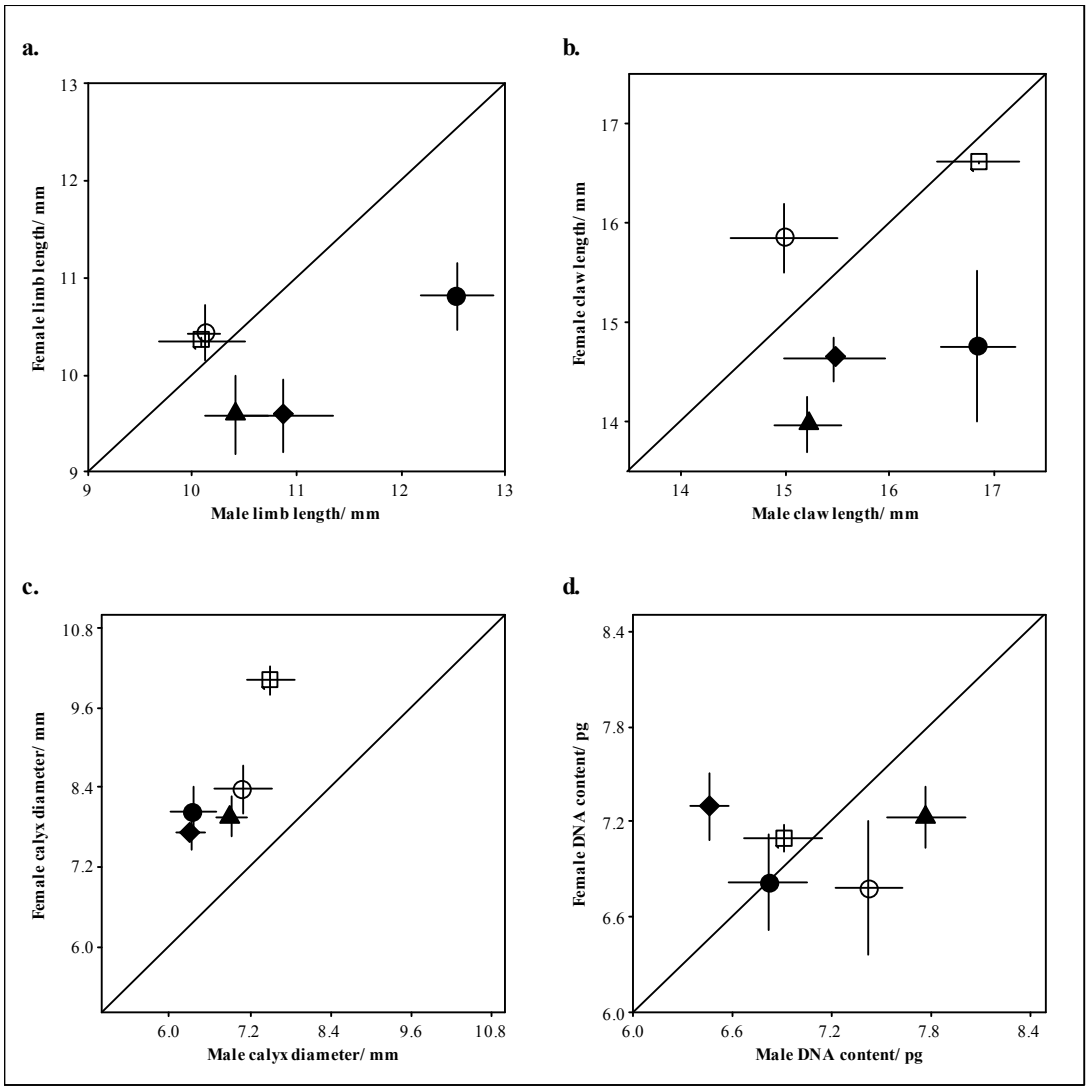


Figure 2.4: Male means plotted against female means for the three floral traits and DNA content in each of the populations studied. Filled circles represent population 15; triangles population 8; diamonds population 14; squares population 13 and open circles population 16. Diagonal lines connect points with equal male and female values. Horizontal and vertical bars show the standard error of the male and female means.

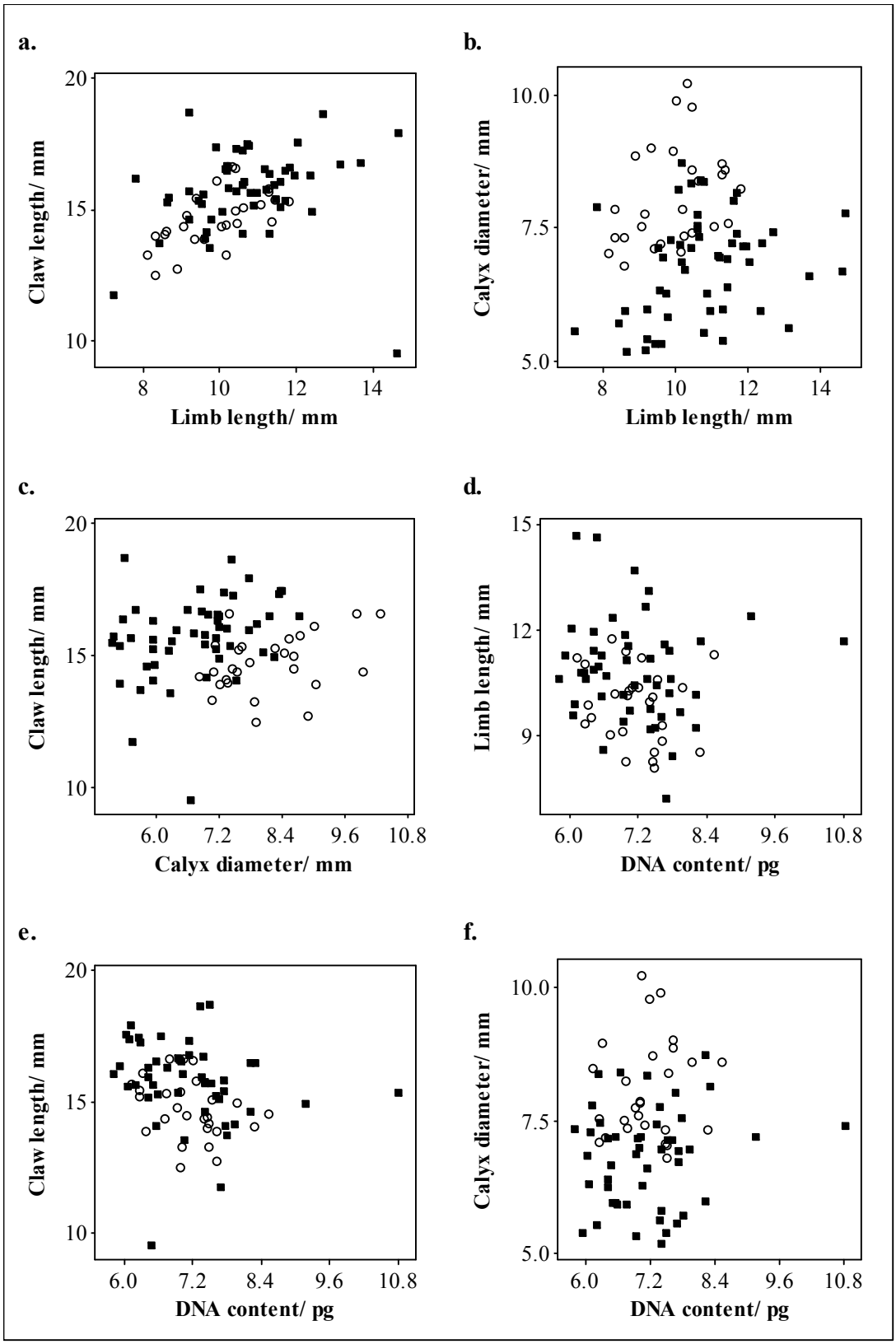


Figure 2.5: Relations between the individual means for the three floral traits and DNA content. Males are represented by filled squares and females by open circles.

Table 2.6: Correlations by populations and sex between the floral traits and DNA content. Numbers in parenthesis indicate sample sizes of females and males respectively.

		Population	13		8		14		15		16		Overall by Sex		Overall	
Trait 1	Trait 2	(F,M)	(2, 8)		(10, 19)		(8, 15)		(3, 6)		(4, 4)		(27,52)			
DNA content	Limb length	Sex	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*
		Male	0.143	0.76	0.164	0.514	-0.35	0.22	0.878	0.021	0.82	0.388	-0.098	0.509		
		Female	-	-	0.118	0.746	-0.27	0.518	-0.105	0.933	-0.218	0.782	-0.183	0.361		
		Overall	0.137	0.725	0.228	0.242	-0.477	0.025	0.444	0.231	-0.214	0.645			-0.109	0.35
DNA content	Claw length	Male	-0.389	0.388	-0.241	0.335	-0.146	0.619	0.584	0.224	0.966	0.166	-0.203	0.167		
		Female	-	-	-0.077	0.833	-0.742	0.035	-0.072	0.954	-0.596	0.404	-0.341	0.081		
		Overall	-0.401	0.285	-0.044	0.822	-0.324	0.142	0.237	0.54	-0.265	0.566			-0.219	0.06
DNA content	Calyx diameter	Male	-0.343	0.452	0.151	0.550	-0.55	0.042	0.319	0.537	0.675	0.528	0.138	0.35		
		Female	-	-	0.341	0.335	0.242	0.563	0.844	0.361	-0.054	0.946	0.119	0.556		
		Overall	-0.105	0.789	0.033	0.867	0.295	0.183	0.277	0.47	-0.387	0.391			0.108	0.358
Limb length	Claw length	Male	0.373	0.363	0.68	0.001	-0.233	0.403	0.292	0.574	0.936	0.064	0.168	0.233		
		Female	-	-	0.780	0.007	0.67	0.072	0.999	0.021	-0.374	0.626	0.605	0.001		
		Overall	-0.295	0.408	0.723	0.000	-0.047	0.831	0.764	0.017	0.327	0.429			0.319	0.004
Limb length	Calyx diameter	Male	-0.287	0.49	0.55	0.016	0.642	0.236	0.224	0.669	0.286	0.714	0.256	0.067		
		Female	-	-	0.401	0.251	0.164	0.698	0.445	0.706	0.01	0.99	0.348	0.075		
		Overall	0.316	0.374	0.281	0.139	0.137	0.534	-0.461	0.212	0.334	0.419			0.074	0.515
Claw length	Calyx diameter	Male	-0.538	0.169	0.715	0.001	0.01	0.396	0.591	0.216	0.321	0.679	0.313	0.024		
		Female	-	-	0.065	0.858	-0.201	0.633	0.475	0.685	-0.581	0.419	0.353	0.071		
		Overall	-0.417	0.231	0.23	0.23	-0.042	0.851	-0.322	0.398	0.341	0.408			0.083	0.467

2.4. Discussion

2.4.1. Germination rates

One substantive concern with the results of the *S. marizii* population survey is the low rate of germination in the survey material, in particular the disparity in germination rates between the five populations. If it is assumed that germination probability is independent of the phenotypic characters being examined then the results should be representative of variation in natural populations. If, however, they are not, then variation in the survey population may be biased compared to natural populations. This is a particular concern, given that seed weight (a potentially significant factor in germination probability) has previously been shown to strongly correlate with genome size in other species (Minelli et al., 1996; Chung et al., 1998). There was also evidence for unequal rates of germination between populations, and between maternal sibships within populations. Neither of these results is particularly surprising, as variation in seed viability or optimal germination conditions might well be expected between populations. Nevertheless, they again suggest that the survey population may not represent a random sample from the *S. marizii* populations that were sampled.

2.4.2. Sex-ratio bias

The results from this study demonstrated a significant male biased sex-ratio in plants which flowered. This male bias appears to be present in almost all of the populations studied.

Sex-ratio bias has been well documented in *S. latifolia*, where a moderate female sex-ratio bias is common in most populations (e.g. Lawrence, 1963; Lyons et al., 1994; Taylor, 1994b, a; Lyons et al., 1995; Taylor, 1996), although variation in both the strength and direction of the sex bias have been documented between families and populations (Lyons et al., 1994; Taylor, 1994b). Further studies of the genetic basis of sex-ratio distortion in *S. latifolia* have shown that it is the action of at least two loci: a Y-linked allele, which is expressed in the paternal plant (and acts to restore equal sex-ratios), and an, as yet, unidentified allele which is expressed in the maternal plant and causes female bias (Taylor, 1994a). These studies have also demonstrated

that the genetic basis of sex-ratio bias may vary between *S. latifolia* populations (Taylor, 1999).

These results suggest that, in contrast to *S. latifolia*, male bias is a more common feature of *S. marizii* populations. However, the survey population was characterised both by low germination rates as well as by low flowering rate and as such it is not clear whether the sex-ratio bias observed in these plants would be representative of sex-ratios in wild populations. Neither is it clear at what point in the development of the plants this bias is introduced. In *S. latifolia*, sex-ratio bias has been shown to exist prior to germination (Lyons et al., 1995), but in this case germination rates were too low for sex-ratios amongst the seed populations to be estimated. In addition, the low rates of flowering, amongst germinated seed, provided another point at which sex bias could be introduced. Indeed, male plants in *S. latifolia* have been shown to flower more readily, and with a greater probability, than females (Lyons et al., 1995). If the same is true for *S. marizii*, then overall sex-ratios (when non-flowering plants are considered) may be much closer to equality than these results suggest.

2.4.3. DNA content variation

Estimates of DNA content in the survey showed considerable variation, ranging from 5.8pg to nearly 11pg, with a standard deviation of 0.8pg. This range is almost certainly an overestimate due to measurement errors associated with the flow-cytometric method (discussed in detail in Chapter 6). Applying the correction method described in that chapter suggests a more realistic range might be 5.8 to 8.1pg, with a standard deviation of 0.5pg, which, nonetheless, still represents an extremely large degree of variation. A striking and surprising result from the analysis of DNA content variation in *S. marizii* is the lack of consistent sexual dimorphism in DNA content, given the extent to which sexual dimorphism in DNA content is present in *S. latifolia*. Even more surprising is the finding of strong evidence that sexual dimorphism in DNA content (when identified at a population level) occurs in different directions in different populations. In *S. latifolia* the sexual dimorphism is due to an enlarged Y-chromosome (Westergaard, 1958; Ciupercescu et al., 1990; Vyskot and Hobza, 2004). This raises the question of why *S. marizii*, a close relative which one might expect to have similar sex-chromosomes, does not appear to share the same effect. These

results suggest that the genetic architecture of the sex-chromosomes may be very different in *S. marizii* from those in *S. latifolia*. Furthermore, they suggest that there is a Y-chromosome polymorphism within *S. marizii*, with some populations showing a relatively larger Y-chromosome, and others showing a relatively larger X-chromosome. It is unclear from these results whether populations lacking sexual dimorphism have approximately equal sized sex-chromosomes, or whether sex-chromosome polymorphism exists within the population. For such a polymorphism to exist within the species would require either a significant degree of genetic isolation between populations or introgression of *S. latifolia* Y-chromosomes in some populations. Either of these interpretations seems plausible. Given the ease with which hybrids form between the two species, it is possible that some of the *S. marizii* populations have been subject to introgressive hybridisation with *S. latifolia* and there also appears to be a considerable degree of morphological differentiation between populations, which is consistent with significant genetic isolation existed between populations. The population level effect on DNA content, detected by the ANOVA is consistent with an explanation based on genetic isolation between populations (although the significant population by sex interaction means that this result must be treated with caution). However, for this to cause sexual dimorphism in DNA content that acts in different directions between populations would require the genetic isolation between populations to be extremely large. This explanation would almost certainly require significant taxonomic sub-structure within the species. An alternative explanation might be that the method of estimating DNA content is affected by heritable variation in an underlying inhibiting factor, which shows variation in sexual dimorphism between populations. Indeed, the levels of sexual dimorphism (where identified) are low compared to those seen in *S. latifolia*. Nevertheless, little evidence for such cytosolic inhibitors was found during an investigation of error associated with the method of DNA content estimation (described in Chapter 6).

2.4.4. Floral trait variation

All the floral traits examined showed an effect of population, as well as a sexual dimorphism, although the strength of these effects varied considerably between traits. The direction of sexual dimorphism identified in calyx diameter is the same as that

previously identified in *S. latifolia*, with females having larger calyces than males and is consistent with the suggestion that large calyces are required in female flowers to accommodate developing ovules and to serve in a protective role. This suggestion is supported by the fact that calyx diameter, in *Silene*, is a trait that can exhibit differences, between the sexes, in ecological significance (Pettersson, 1991; Wright, 2000; Wright and Meagher, 2004).

However, in contrast to patterns previously identified in *S. latifolia*, sexual dimorphism in limb length and claw length act in the opposite direction, with *S. marizii* males having longer limbs and claws than females. The pattern of sexual dimorphism identified here seems more consistent with Bateman's principle and the assumed functional significance of the three floral traits. It is also more consistent with previous surveys of floral displays in dioecious plants (Bell, 1985; Delph et al., 1996; Humeau et al., 2003; Miller and Venable, 2003), which have identified larger flowers in male plants to be the most common pattern of sexual dimorphism. In *S. latifolia*, this apparent discrepancy has been explained by the overall floral display, with males having a greater number of flowers than females despite the smaller size of each flower (Lloyd and Webb, 1977; Meagher, 1992; Steven et al., 2007). Whilst flower number was not measured in this glasshouse study, qualitative observations suggest that males of *S. marizii* also produce many more flowers than females. It has been demonstrated that strong genetic correlations with calyx diameter may be responsible for the associated direction of sexual dimorphism in other floral traits rather than independent selection (Meagher, 1992, 1999; Delph et al., 2002; Delph et al., 2004). These results suggest that the genetic correlations among floral traits may not be as strong in *S. marizii*, or that the historical intensity of sex-specific selection has not been as strong as appears to have been the case for *S. latifolia* and thus there is more potential for independent sex-specific evolution in calyx size and petal size. Hence, there may be significant differences in the genetic basis of floral variation between the two species.

2.4.5. Correlations between floral traits and DNA content

The lack of strong correlations between DNA content and floral characters contrasts with such correlations that have previously been reported in *S. latifolia* (Meagher and

Costich, 1994, 2004). There is some evidence for a negative correlation between claw length and DNA content, a correlation that appears to be strongest in female plants. Nevertheless, taken together with the sexual dimorphism identified in the floral traits, the lack of correlations with DNA content suggests that the genetic basis of floral trait determination and sexual dimorphism in particular may be very different in *S. marizii* from that in *S. latifolia*.

2.4.6. Correlations between floral traits

The only strong correlation detected among the three floral traits was between limb length and claw length. This correlation is unsurprising given that both characters are, to a large extent, part of the broader character of petal size. As such it would be no surprise to find that there is a strong genetic correlation between the two, as there is in *S. latifolia* (Meagher, 1999; Delph et al., 2004). The lack of correlation between calyx diameter and either of the other two characters contrasts with relationships previously reported in *S. latifolia* (Meagher, 1992). Again this supports the idea that patterns of genetic correlations are different in *S. marizii* and that the potential for independent sex-specific evolution exists between calyx diameter and petal size within *S. marizii*.

2.4.7. Summary

The survey of DNA content and floral variation in *S. marizii* demonstrates that the pattern of sexual dimorphism in DNA content, which has been reported, and is well established, in *S. latifolia*, is not a ubiquitous characteristic of the section *Elisanthe*. Furthermore, these results demonstrate that sexual dimorphism in DNA content can occur in different directions within a single species. Although it has not been conclusively established that a Y-chromosome polymorphism is responsible for these sexual dimorphisms, it is hard to envisage another mechanism which could be responsible. As such, it would be extremely useful for further investigation into the nature of DNA content variation in *S. marizii* to adopt a cytogenetic approach in order to directly observe sex-chromosome heteromorphy within the species. If it were established that a Y-chromosome polymorphism is indeed present within the species, then the mechanism by which such a polymorphism is maintained within a single

species would warrant further investigation. The most likely candidates for such a mechanism would be either Y-chromosome introgression from *S. latifolia* in certain populations, or historical isolation between populations giving rise to substantial taxonomic sub-structure within *S. marizii*. The results also raise the possibility that an enlarged Y-chromosome (relative to the X) may not be a constant feature of dioecious *Silene* species.

Another general result from this survey is the lack of relationship between calyx diameter and petal measurements, both in the direction of sexual dimorphism and phenotypic correlations. This seems to suggest the potential for independent sex-specific evolution between these characters. This contrasts with previous results from *S. latifolia*. The genetic correlations between calyx size and petal size are well established in *S. latifolia* suggesting that it is differences in the genetic control of floral development between the two species, rather than variation in patterns of sex-specific selection, that are responsible for this difference. A quantitative genetic approach would be useful to confirm that the strength of genetic correlations between floral traits are, indeed, weaker in *S. marizii*.

Chapter 3 The dynamics of DNA content variation in interspecific hybrids of *S. latifolia* and *S. marizii*

3.1. Introduction

Large differences exist in DNA content between *Silene* species of the section *Elisanthe*. Furthermore, these differences in DNA content are correlated with variation in flower size (Meagher and Costich, 2004). The use of hybrids between two closely related members of the section (*S. latifolia* and *S. marizii*) may provide a valuable resource with which to study the genetic basis of this relationship. Hybrids between these species also offer the opportunity to study the role which DNA content might play in genetic isolation between the two species. The work described in this chapter uses an F2 crossing design to examine rates of germination, patterns of DNA content variation and floral variation in two generations of *S. marizii*/ *S. latifolia* hybrids.

3.1.1. Germination rates

The use of *S. latifolia* and *S. marizii* as a model system to study the nature of intraspecific differences in DNA content variation and the genomic basis of correlated variation in flower size, relies both on the presence of interspecific variation in DNA content and on the ability of the two species to produce viable seed following hybridisation. A predominant feature of the previous work with *S. marizii* (Chapter 2) and hybrids of *S. marizii* and *S. latifolia* has been very low rates of germination,

accompanied by low rates of flowering amongst germinated seedlings. This was particularly so during attempts to grow a population of F2 hybrids for a QTL mapping study (which had been part of the original research plan). Low germination rates, particularly if non random may have important consequences for the interpretation and utility of both phenotypic surveys and genetic mapping studies. This is a particular worry, given that seed weight (a potentially significant factor in germination rate) has previously been shown to strongly correlate with genome size in other species such as soybean (*Glycine max*) (Chung et al., 1998) and faba bean (*Vicea faba*) (Minelli et al., 1996), where germination power was also shown to correlate with DNA content. As such, variability in germination rates has the potential to cause significant bias in a hybrid population with regards to DNA content (the trait of interest).

It is not possible to measure DNA content in ungerminated seed, and as such, it is impossible to directly compare mean values for this trait between germinated and ungerminated seed. However, if the bias is introduced through an effect on seed weight, this link might be established by examining the effect of seed weight on germination rates, and the effect of seed weight on the traits of interest.

3.1.2. DNA content variation in interspecific hybrids

A number of previous studies have examined patterns of DNA content variation in interspecific hybrids between parents with large differences in DNA content. In general, such studies have found that F1 hybrids show DNA content that is intermediate to that of the two parents (e.g. Barre et al., 1998; Morgan et al., 1998; Thibault, 1998; Sisko et al., 2003; Vaio et al., 2007). However, a number of studies have reported considerable variation in DNA content amongst F1 hybrids (Price et al., 1983, 1985; Rayburn et al., 1993). Price et al (1983) in particular noted that considerable variation in the estimated DNA content in the F1 hybrids of *Microseris* appeared to be heritable. This was interpreted as evidence that DNA content might be unstable during hybridisation, and susceptible to either deletion or amplification. A more recent study of natural interspecific hybrids has reported similar distortions in DNA content relative to the parents (Bures et al., 2004), with natural hybrids between *Cirisum* species showing decreased DNA content compared to the mean of their

parent species. As such, a further aim of this study was to examine segregation of DNA content amongst *S. marizii*/ *S. latifolia* hybrids. In particular, patterns of DNA content variation within and between hybrid populations were examined in order to determine whether between species differences in DNA content are autosomal in origin or due to differences in sex chromosomes. If the former were true, then one would expect mean DNA content to be independent of the direction of the cross, whilst in the later case mean DNA content in hybrid populations will depend on the direction of the cross, with the presence or absence of sexual dimorphism depending on the relative size of each of the sex chromosomes from each species.

3.1.3. Correlations between DNA content and flower size

A further question that can be studied using interspecific hybrids is whether or not the genetic basis of the interspecific correlation between DNA content and flower size, which is present within *S. latifolia*, has a similar genetic basis to the interspecific relationship between these traits. If so, one would expect to find a strong relationship between these traits in F2 hybrids. As such, the final objective of this study was to test whether a relationship between DNA content and flower size is present in hybrids between the two species.

3.2. Methods

3.2.1. Origin of seed

Fifty seeds from each of 7 female plants were selected. These seeds represented samples from each of the three generations of an F2 crossing design. These crosses and the populations from which the seed samples were drawn are shown schematically in Figure 3.1. The parental seed populations P1 and P2 were generated from intraspecific crosses within *S. latifolia* and *S. marizii,* respectively. Two other samples were F1 hybrid seed that were generated by reciprocal interspecific crosses between plants grown from the parental seed. The final three seed samples represented F2 hybrids generated by crosses made within each of the two F1

populations. Interspecific crosses were made by the physical transfer of pollen from a paternal flower to a maternal flower. Developing flowers on the maternal plant were isolated in sealed bags, prior and subsequent to crossing, to prevent contamination by other pollen. In all cases control flowers were left unpollinated to test the effectiveness of the method of isolation.

Each seed used in the study was weighed to the nearest 0.1mg prior to planting.

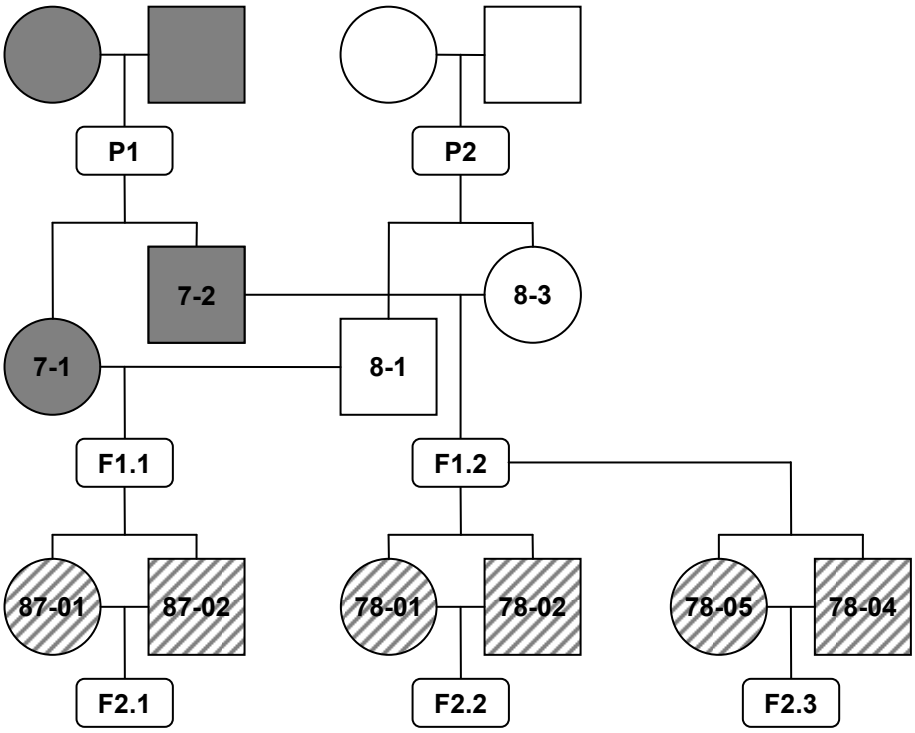


Figure 3.1: Schematic diagram showing the origin of seed material used in the study. Circles represent female individuals; squares represent males. White symbols indicate *S. marizii* individuals; grey symbols indicate *S. latifolia*; and alternating stripes indicate an *S marizii*/ *S latifolia* hybrid. Rectangles show the populations of seeds from which individuals for further crosses were generated, and from which seeds for the study were drawn.

3.2.2. Cultivation of plant material and phenotypic measurements

Seeds were planted in a randomised design under germination conditions identical to those described in Chapter 2. Germination trays were examined daily, and the date of germination, where appropriate, was noted for each seed. When germinated seedlings reached a sufficient size, flow-cytometric estimates of DNA content were made using the procedure described in Chapter 2. For plants that survived to floral maturity, measurements of floral traits were made (as described in Chapter 2).

3.2.3. Statistical analysis

Due to the low rates of germination in hybrid populations, especially in the F1 populations, DNA content estimates and flower size measurements from previously generated plant material (from the same seed samples) were included in the analysis of sex-ratio bias, and the analysis of phenotypic variation within hybrid populations.

Sex-ratios in flowering plants within each population were tested against the expectation of equal numbers of males and females using a chi-squared goodness of fit test.

To examine the effect of seed weight on germination probability, differences in the mean seed weight of germinated and ungerminated seed from each population were examined using a two sample *t*-test.

Segregation of DNA content amongst hybrids was examined. Homogeneity of variance in DNA content among hybrid populations was tested using Levene's test. The effect of parental cross direction and population (nested within parental cross direction) was tested using an ANOVA examining the effects of cross direction, generation (F1 or F2) and population nested within cross and generation. Comparisons between the means for each cross, generation and population were made using Tukey's method.

Relationships between the variables examined were tested using Pearson's correlation coefficients. In addition to the overall correlation coefficients, coefficients within species and sexes were also calculated.

3.3. Results

3.3.1. Germination rates

Germination was entirely absent in three of the populations examined. These were: P1, F1.1 and F1.2. Germination rates were highly variable in the other populations,

ranging from 12% to 60% (Table 3.1). Amongst the F2 populations germination rates were highest in population F2.1 and F2.2, but substantially lower in population F2.3.

Rates of flowering amongst germinated seed were also highly variable between populations, ranging from 0% in population F2.3 to 48% in population F2.2 (Table 3.1).

There was evidence for an excess of females compared to males in flowering plants from population F1.1. Whilst there was no evidence for a sex-ratio bias in any of the other populations (Table 3.1), the sex-ratio in population F2.1 was comparably low, but not statistically significant due to the very low sample size.

Table 3.1: Summary of germination and flowering rates amongst the seed populations tested. The total numbers of flowering plants (from all sources) are shown along with the sex-ratios and their statistical significance. In populations where rates of germination or flowering were particularly low, sex-ratios were taken from previously germinated seed.

Population	Seed	Geminated (rate)	Flowered (rate)	N	Sex-ratio (males/ total)	χ^2	*p*
P1	50	0 (0.00)	-	0	-	-	-
P2	50	15 (0.30)	3 (0.20)	3	0.33	0.33	0.564
F1.1	50	0 (0.00)	-	14	0.14	7.14	0.008
F1.2	50	0 (0.00)	-	9	0.56	0.11	0.739
F2.1	50	30 (0.60)	5 (0.17)	5	0.20	1.80	0.180
F2.2	50	29 (0.58)	14 (0.48)	15	0.53	0.07	0.796
F2.3	50	6 (0.12	0 (0.00)	7	0.43	0.14	0.705
Total	350	80 (0.23)	22 (0.28)	56	0.37	0.53	0.465

3.3.2. Effect of seed weight on germination of F2 seed

There was strong evidence to suggest that germinated seed were heavier than ungerminated seed in two of the F2 populations (Figure 3.2). There was also strong evidence for a negative correlation between seed weight and germination time, a relationship that was present over all, as well as within each of the populations (Table 3.2, Figure 3.3).

Table 3.2: Correlation coefficients and associated *p*-values between seed weight and germination time. Correlations within and over all populations are shown. Numbers in parentheses indicate the sample size within each population.

Population	*r*	*P*
F2-1 (30)	-0.44	0.015
F2-2 (29)	-0.404	0.030
F2-3 (6)	-0.509	0.302
Overall	-0.546	0.000

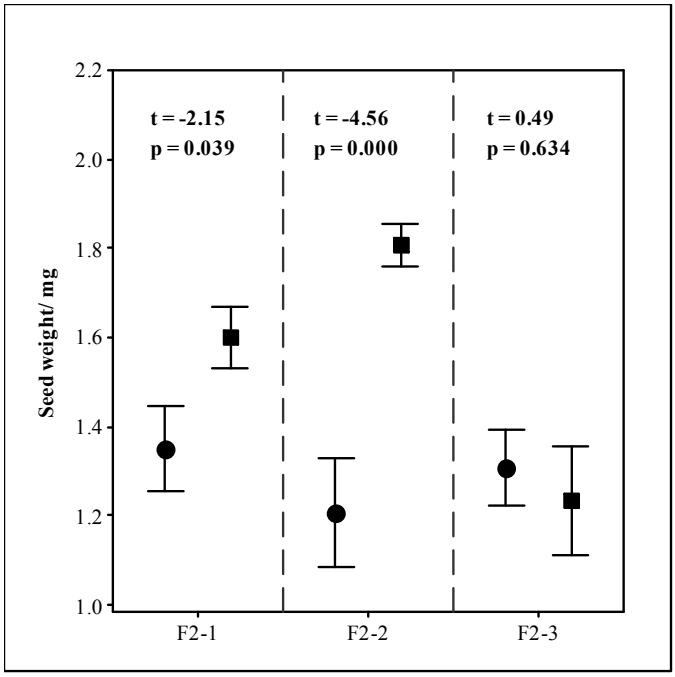


Figure 3.2: Mean seed weights of germinated (squares) and ungerminated seeds (circles) in each of the three F2 populations. Vertical bars represent the standard error in the mean. The results of a two tailed *t*-test examining the difference in the mean of the germinated and ungerminated seed from each population are also shown.

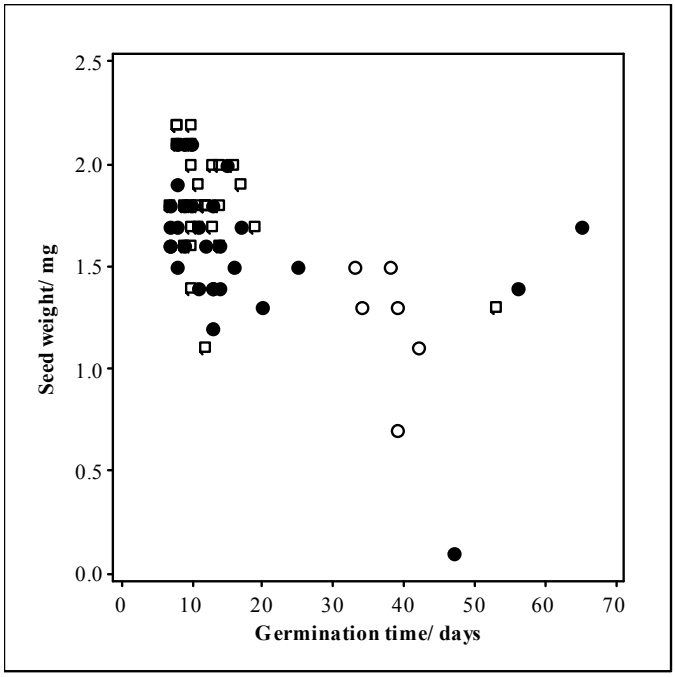


Figure 3.3: Relationship between seed weight and germination time in the three F2 populations. Filled symbols represent population F2.1, crossed symbols population F2.2 and open symbols population F2.3.

3.3.3. Correlations between seed weight and phenotype in F2 populations

There was no evidence for a relationship between seed weight and DNA content in any of the F2 populations studied, or across all populations. Neither was there any evidence for a relationship between seed weight and any of the three floral traits (Table 3.3, Figure 3.4).

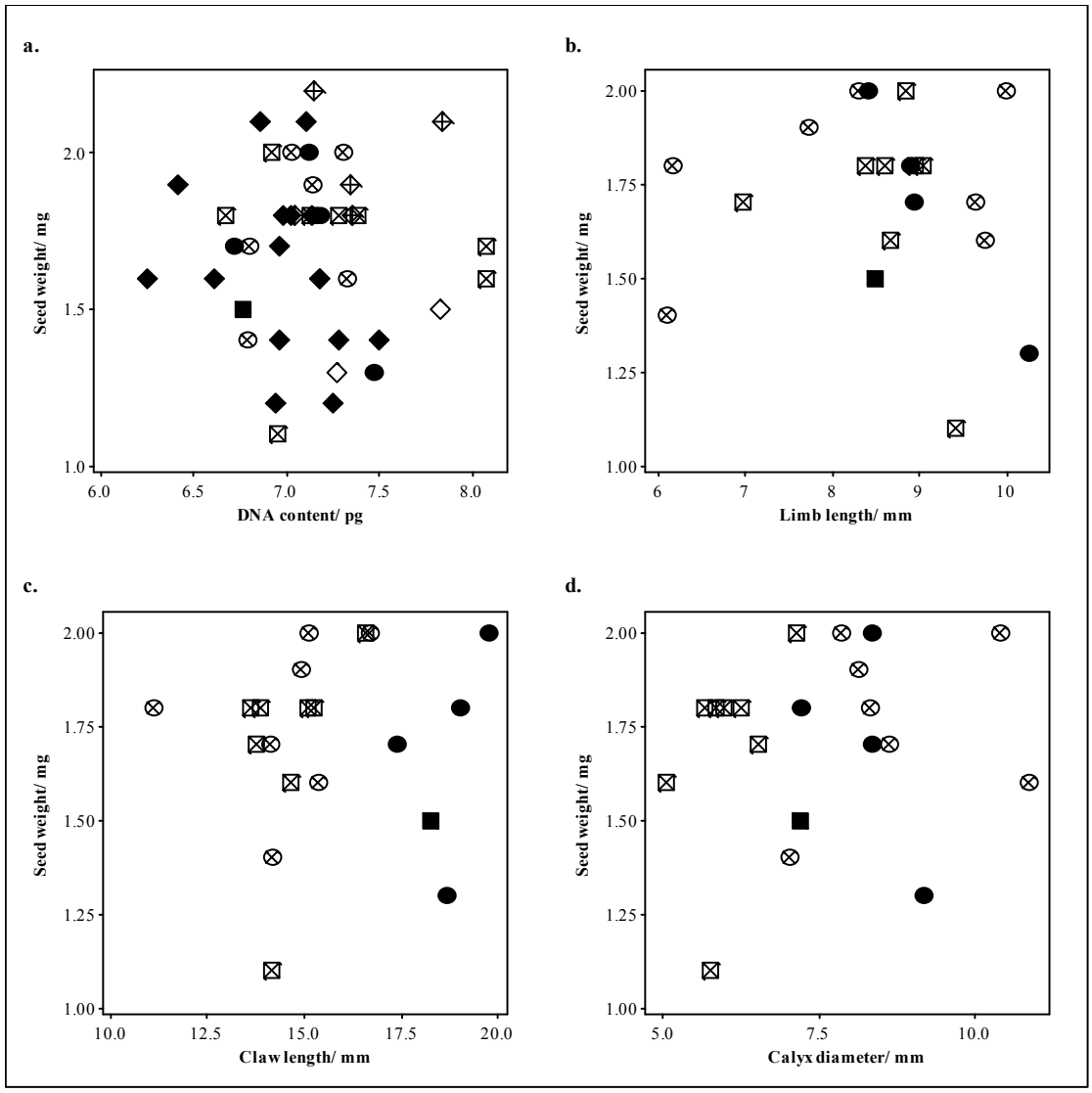


Figure 3.4: Relationship between seed weight and a: DNA content; b: limb length; c: claw length; and d: Calyx diameter. Filled symbols represent population F2.1, crossed symbols population F2.2 and open symbols population F2.3. Males are represented by squares, females by circles and non-flowering plants by diamonds.

Table 3.3: Correlation coefficients and *p*-values between seed weight and various traits. Correlations within and over all populations are shown. Numbers in parentheses indicate the number of plants with DNA content estimates and the number of plants with floral measurements respectively.

Population	F2.1 (21, 5)		F2.2 (19, 14)		Overall	
Trait	*r*	*p*	*r*	*p*	*r*	*p*
DNA content	0.29	0.235	0.12	0.640	-0.02	0.892
Limb length	-0.74	0.153	0.08	0.797	-0.08	0.743
Claw length	0.45	0.452	0.29	0.317	0.09	0.726
Calyx diameter	-0.30	0.626	0.27	0.350	0.15	0.534

3.3.4. DNA content variation in hybrid populations

The variance of DNA content estimates tended to be greater in F1 populations than in F2 populations (Table 3.4), although this difference was not significant (Levene's test statistic= 0.71, p=0.586).

Table 3.4: Mean and standard deviation of DNA content in each of the hybrid populations.

Population	N	Mean	Std Dev
F1-1	8	6.52	0.691
F1-2	7	6.75	0.502
F2-1	21	7.00	0.313
F2-2	19	7.24	0.399
F2-3	10	7.76	0.577

Table 3.5: ANOVA testing for the effects of cross direction, generation, population (nested within cross and generation) and cross by generation interaction on DNA content in hybrid populations.

Source	Degrees of freedom	Sum of squares	Adjusted sum of squares	Adjusted mean square	*F*-ratio	*p*
Cross	1	2.89	1.51	1.51	7.17	0.010
Generation	1	3.76	4.27	4.27	20.21	<0.001
Population (Cross Generation)	1	1.61	1.71	1.71	8.10	0.006
Cross x Generation	1	0.21	0.21	0.21	1.00	0.321
Error	60	12.67	12.67	0.21		
Total	64	21.15				

There was strong evidence that the direction of the cross affected DNA content. DNA content was significantly less when *S. marizii* was the male parent (difference in the mean = 0.366 pg, t=3.22, p=0.002). There was also very strong evidence that generation had a strong effect on DNA content, with the F2 generation showing significantly higher DNA content than the F1 (F2-F1= 0.62 pg, t=-3.63, p=0.002).

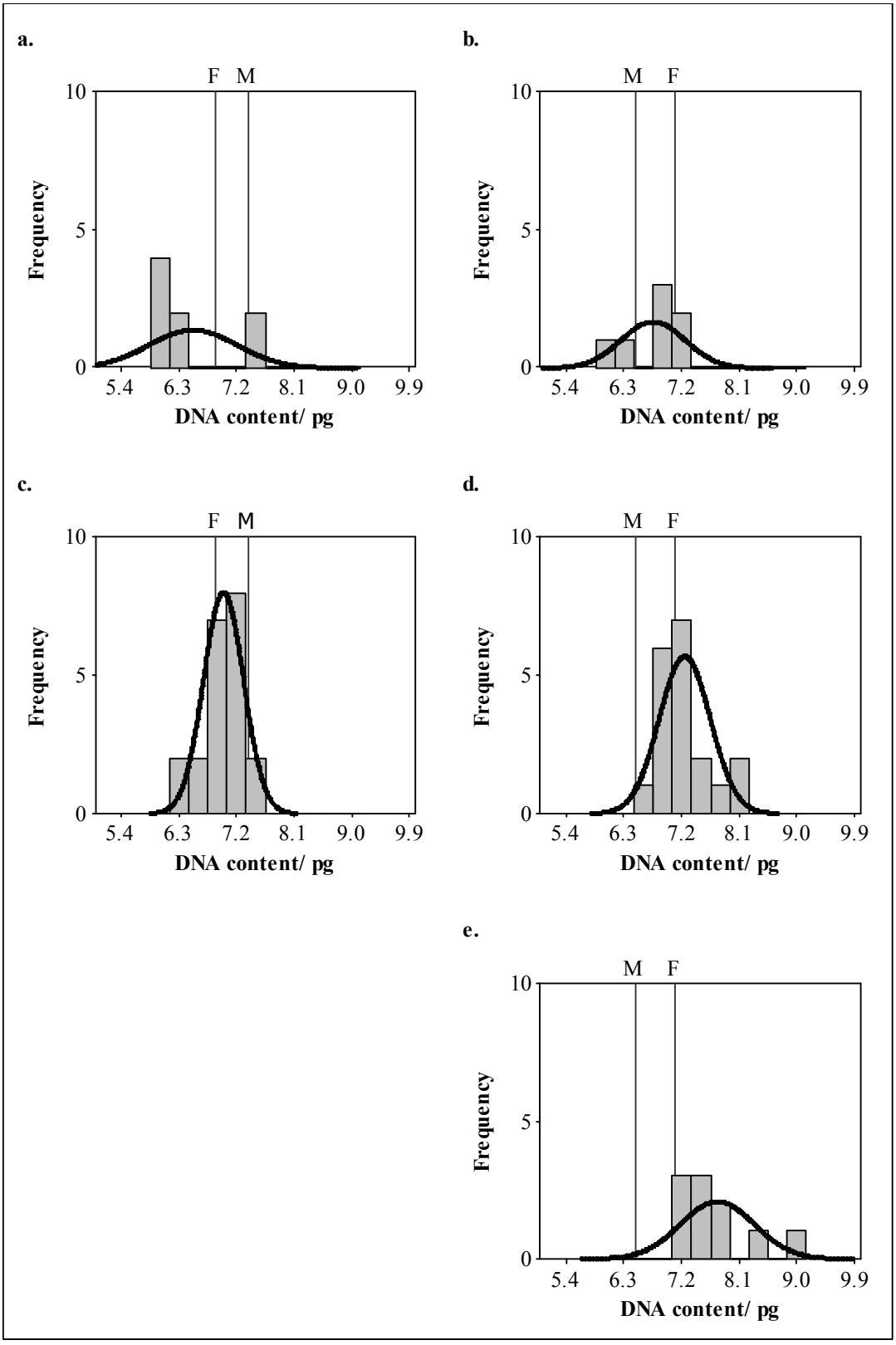


Figure 3.5: Frequency distributions of DNA content estimates within each of the hybrid populations, with fitted normal distributions. Vertical lines show the mean DNA content estimates of the parents used to generate each of the crosses (F indicates the maternal parent and M the paternal parent). Panel a represents population F1.1; b: F1.2; c: F2.1; d: F2.2; and e: F2.3.

There was no evidence for a significant interaction between the direction of the parental population and generation. Population (nested within cross and generation) also showed strong evidence for an effect on DNA content. Comparisons of population means revealed that there was a significant difference between the means of population F2.2 and F2.3 (difference= 0.51pg, t= 2.85, p= 0.046). In every hybrid population studied, the range of DNA content variation exceeded the interval between the high and low estimates of parental DNA content (Figure 3.5). In population F1.1, F2.2 and F2.3, the population mean DNA contents were not intermediate to those of the parents.

3.3.5. Phenotypic correlations in F2 hybrids populations

There was no consistent evidence for a relationship between any of the traits over all sexes and populations. There was a general trend towards negative correlation coefficients between DNA content and claw length, although this was not statistically significant (Table 3.6, Figure 3.6)

Table 3.6: Correlation coefficients between phenotypic traits in F2 populations by population and sex. Numbers in parentheses represent the numbers of males and females respectively.

		Population (M,F)	**F2.1 (1,4)**		**F2.2 (8,7)**		**F2.3 (3,4)**		**Overall (12,15)**	
Trait 1	**Trait 2**	**Sex**	***r***	***p***	***r***	***p***	***r***	***p***	***r***	***p***
DNA content	Limb length	Male	-	-	-0.702	0.052	-	-	-0.058	0.865
		Female	0.626	0.374	0.289	0.578	0.077	0.923	0.409	0.147
		Overall	0.69	0.197	-0.221	0.447	-0.121	0.819	0.182	0.384
DNA content	Claw length	Male	-	-	-0.179	0.672	-	-	-0.399	0.224
		Female	0.591	0.409	0.466	0.352	0.417	0.583	-0.303	0.293
		Overall	0.608	0.276	-0.088	0.764	-0.407	0.423	-0.323	0.116
DNA content	Calyx diameter	Male	-	-	-0.238	0.571	-	-	-0.339	0.308
		Female	0.298	0.702	0.416	0.412	-0.664	0.336	0.057	0.846
		Overall	0.498	0.393	-0.22	0.449	-0.081	0.879	-0.051	0.809
Limb length	Claw length	Male	-	-	0.306	0.461	0.999	0.006	0.044	0.892
		Female	-0.251	0.749	0.709	0.075	-0.238	0.762	0.194	0.489
		Overall	-0.144	0.817	0.614	0.015	0.14	0.765	0.137	0.495
Limb length	Calyx diameter	Male	-	-	-0.26	0.534	0.674	0.529	-0.001	0.998
		Female	0.614	0.386	0.782	0.038	0.182	0.818	0.532	0.041
		Overall	0.682	0.205	0.212	0.449	-0.098	0.834	0.152	0.448
Claw length	Calyx diameter	Male	-	-	0.557	0.152	0.666	0.536	0.554	0.062
		Female	-0.155	0.845	0.472	0.285	-0.956	0.044	-0.142	0.615
		Overall	-0.001	0.999	0.239	0.39	-0.845	0.017	0.044	0.828

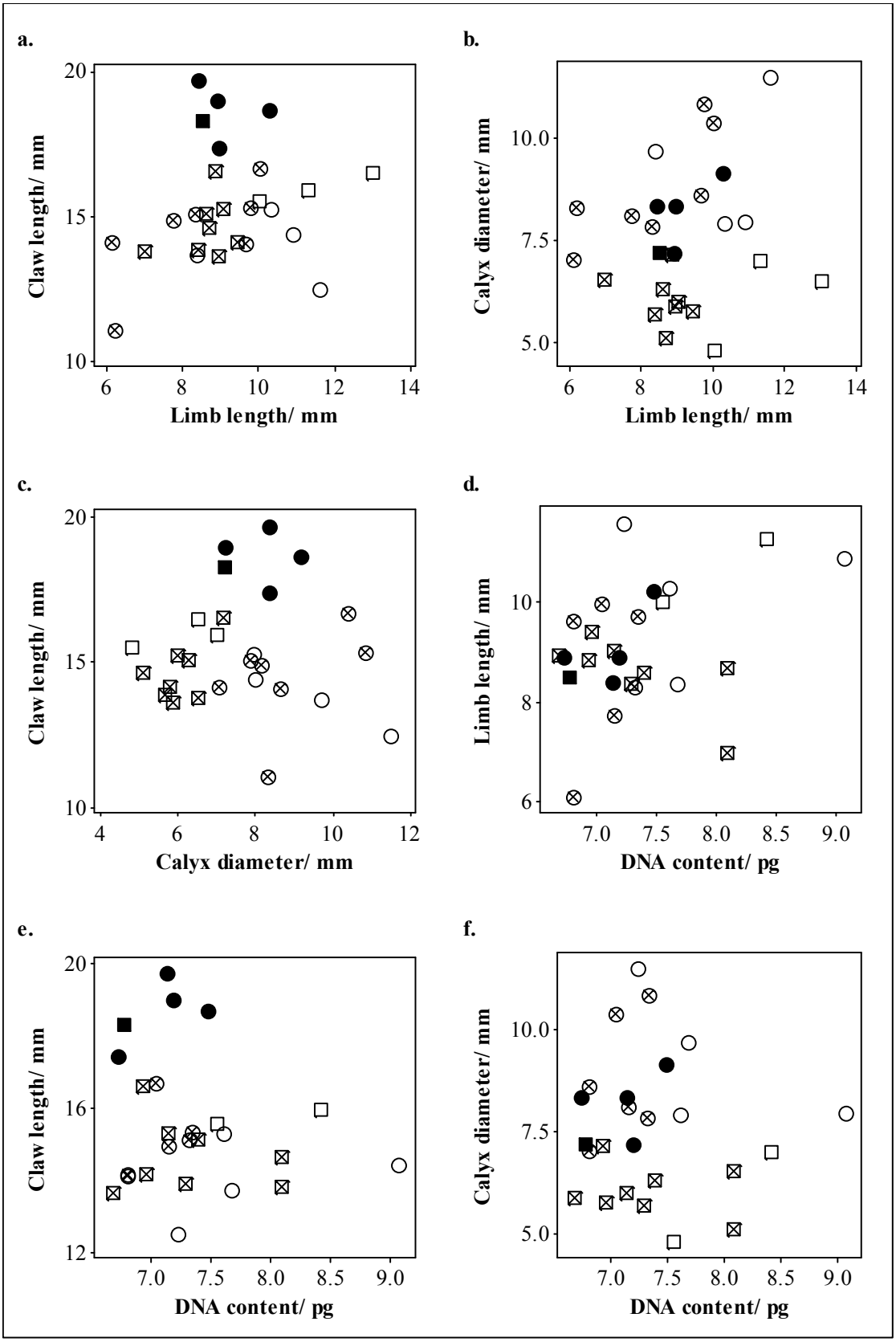


Figure 3.6: Relationships between phenotypic traits in populations of F2 hybrids. Filled symbols represent population F2.1, crossed symbols population F2.2 and open symbols population F2.3. Males are indicated by squares; females by circles.

3.4. Discussion

3.4.1. Germination rates

The extremely low germination rates found in the two parental seed populations are surprising. Nevertheless, seed from these populations are substantially (approximately four years) older than seed from the other populations and it seems likely that this result simply reflects the age of the seed. The complete absence of germination in both of the F1 seed populations is more surprising. The seed are not substantially older than the F2 seed, and were stored under identical conditions.

Germination rates in the F2 seed populations were highly variable. This result is interesting, and suggests that seed viability may vary between crosses, even between crosses within a single F1 population. It is unclear from these results how such variation in seed viability might have been induced.

3.4.2. Hybrid sex-ratios

Evidence for a sex bias was found in a single F1 population (F1.1), where there was a strong female bias and there was the suggestion of a similar sex bias in the corresponding F2 population (F2.1) although this was not significant. However, the sample size was low. This is an interesting result, and is consistent with Haldane's rule (Haldane, 1922; Coyne, 1985); males being the heterogametic sex. It is also consistent with the proposed genetic control of sex-ratio bias in *S. latifolia*, where an unidentified female-expressed distorter allele is thought to cause female bias, with a Y-linked allele acting to restore equal sex-ratios (Taylor, 1994b). In both population F1.1 and F2.1 the *S. marizii* Y-chromosome is present in all males. This result suggests both that *S. marizii* males lack restorer alleles found in *S. latifolia* males, and that the effect of the distorter allele in females may be strong in the absence of the restorer allele. Indeed, sex-ratios in population F2.1 (where the *S. marizii* Y-chromosome also segregates) appear to show a similar bias towards females, although with few flowering plants the power to detect statistically significant effects is low. Nevertheless these results suggest that sex-ratio distortion may be a significant feature of *S. latifolia* / *S. marizii* hybrids. The lack of a male sex bias in populations of the

reciprocal cross (F1.2, F2.2 and F2.3) supports the suggestion that *S. marizii* lacks this mechanism of sex-ratio distortion, and that the Y-linked restorer allele does not influence sex-ratios independently of the distorter allele. This contrasts with previous studies of *Silene* hybrids (*S. latifolia* and *S. dioica*), where female biased sex-ratios were present in hybrids, regardless of the direction of the cross (Taylor, 1994a). Nevertheless, Taylor (1999) demonstrated that substantial variation exists, within populations, in the alleles that are responsible for sex-ratio bias in *S. latifolia* and the absence of sex-ratio bias may simply reflect the genotype of the individuals selected for theses crosses.

3.4.3. Effect of seed weight on germination of F2 seed

There is strong evidence to suggest that seed weight is a significant determinant of germination probability within F2 hybrids. This is demonstrated both by the fact that seeds which germinate are, in general, heavier than seeds that do not, and by the fact that the time until germination is negatively correlated with seed weight. This result does not rule out another factor besides seed weight that causes low germination in hybrid seed, but it does suggest that a great proportion of the variation in germination probability can be attributed to seed weight. As such, there is some justification for using correlations between seed weight and DNA content or flower size to answer the question of whether low germination rates in hybrid populations are likely to cause a bias with regard to these traits.

3.4.4. Correlations between seed weight and phenotype in F2 populations

The lack of evidence for any relationship between DNA content or floral dimensions and germination time suggests that germination probability is likely be independent of these characters and that hybrid populations are unlikely to be biased with regard to these traits. Nevertheless, the apparent bias in DNA content in hybrid populations (discussed in more detail below) casts some doubt on this conclusion.

3.4.5. DNA content variation in hybrid populations

The strong effect of the direction of the parental cross on DNA content amongst hybrid populations is consistent with a smaller *S. marizii* Y-chromosome relative to that found in *S. latifolia*. However, there is also evidence for a significant female biased sex-ratio in populations derived from the cross that uses *S. marizii* as the male parent and, as such, this disparity would also be consistent with a sexual dimorphism in DNA content in these plants. Under this interpretation, differences in sex-chromosome heteromorphy between the two species would be more likely to be caused by an enlarged X-chromosome in *S. marizii* relative to that in *S. latifolia*. Unfortunately, the extreme sex-ratio bias in these populations prevented a reliable test for sexual dimorphism. Nevertheless, if differences in the size of the Y-chromosome were responsible for differences in DNA content between the two crosses, then one would also expect to see a sexual dimorphism in DNA content within populations in which the *S. latifolia* Y-chromosome was segregating. When a *t*-test is performed to test for this effect, there is only a very small difference in the means for each sex (0.06 pg), a difference that is not significant (t=-0.24, p=0.814). This suggests that differences in sex-chromosome heteromorphy between the two species can be better explained in terms of the relative sizes of the X-chromosomes in the two species. The significant difference between the mean of population F2.2 and F2.3 is surprising given that both populations derive from the same parental cross. This difference suggests either that there is considerable variation in DNA content amongst the F1 (as has also been reported in *Microseris* (Price et al., 1983)) or that segregation of DNA is distorted in one or other of the populations. Indeed, there is strong evidence that DNA content in F2 hybrids is significantly greater than that found in F1 hybrids. However, given the range of DNA content estimates in F1 hybrids, and the small number of populations, either explanation is consistent with the data.

It is notable that in neither case are the variances in DNA content significantly different between the F1 generation and the F2 generation of each cross. One would naturally expect F1 hybrids to cluster around the parental midpoint, with segregating DNA content loci in the F2 causing greater variance in DNA content. However, other work in this thesis (Chapter 6) has shown that significant error may be present in the method of DNA content estimation and the high variance in each of the hybrid

populations may simply reflect a high measurement error relative to the true variance in DNA content. Nevertheless, a similar pattern of variation to that seen here has previously been reported in *Microseris* (Price et al., 1983), where it was interpreted as evidence that elements representing DNA content differences were unstable and capable of deletion or amplification during hybridisation. Together with the suggestion that there is an upward trend in DNA content in successive generations of the F2 crosses (see above), these results indicate that DNA content in hybrid populations may be biased compared to those of the parents, although the point at which this bias is introduced is unclear from these data. Nevertheless, these results may well reflect mechanisms of reproductive isolation between the two species that almost certainly exist, given their overlapping distributions and the ease with which hybrids are formed. In particular, they raise the possibility that DNA content may play a role in maintaining the species identity rather than strong selection associated with ecological isolation, as has been reported within other members of the section (Goulson and Jerrim, 1997).

3.4.6. Phenotypic correlations in F2 hybrids

The lack of significant correlations between DNA content and any of the three floral traits is surprising given the magnitude of DNA content and flower size differences between the parents. This result suggests that many of the genetic loci that are responsible for interspecific variation in flower size are independent of DNA content. This does not contradict the strong correlations reported in *S. latifolia*, and is consistent with the suggestion that the genetic control of these characters may be very different in *S. marizii* (Chapter 2). The lack of significant correlations between the three measures of flower size also suggests that loci affecting interspecific differences in flower size may be very different from those affecting intraspecific variation. Nevertheless, the suggestion that hybrid populations are biased in DNA content relative to the parents casts some doubt over the ability of detecting such relationships in hybrid populations.

3.4.7. Summary

These results confirm that low rates of germination and flowering in hybrids of *S. marizii* and *S. latifolia* are a consistent feature, and that there is some evidence to suggest that hybrid populations may be biased with regard to at least one of the traits of interest. Indeed, the patterns of DNA content segregation in hybrids are surprising and suggest that a substantial bias in DNA content in hybrids may be introduced during the process of hybridisation. This result is intriguing, and warrants further investigation. In particular, further studies with greater sample sizes (both in terms of populations and individuals) may reveal whether high variability in the DNA content of F1 seed is genuine or simply an experimental artefact, and whether DNA content in hybrids is biased with respect to the parents. If such findings were confirmed, then this would have significant repercussions on the utility of the *S. marizii*/ *S. latifolia* system, but the mechanisms by which such a bias arose would be of considerable interest in their own right, particularly the role that DNA content might play in the maintenance of species boundaries.

Sex-ratio bias also appears to be a feature of hybrids, but only when *S. marizii* is used as the male parent. This suggests that genetic mechanisms of sex-ratio distortion, previously identified in other *Silene* species, may not be present in *S. marizii*. As such, hybrids between these two species may provide a useful system with which to study the evolution of sex-ratio distorters.

Another interesting result of this study is the lack of correlation between DNA content and flower size in F2 hybrids. This suggest that loci affecting flower size and DNA content differences between the two species may be different in nature to those that cause differences in these traits in *S. latifolia*. Hence, studying such hybrids may not greatly increase understanding of the mechanism by which DNA content may influence phenotype, although biases in DNA content amongst hybrid populations may also explain these results.

The results suggest that a significant proportion of the variation in DNA content between the two species may be due to differences in sex-chromosome size and that these differences may be attributable to variation in the size of the X-chromosome.

This is an interesting result, but requires verification. In particular, evidence for a sexual dimorphism in DNA content in hybrids derived from crosses using *S. marizii* as the paternal parent would provide support for this hypothesis.

Chapter 4 Estimating the contribution of Ty1-*copia* class retrotransposons to DNA content using site-specific amplification polymorphism

4.1. Introduction

4.1.1. LTR retrotransposons

Retrotransposons are the most common form of transposable element found in eukaryote genomes and are characterised by a transposition involving an RNA intermediate (Bingham and Zachar, 1989). They are classified into either Long Terminal Repeat (LTR) or non-LTR types based on the presence or absence of long sections of repeated sequence at each end of the retrotransposon (the long terminal repeat) (Boeke and Corres, 1989). LTR retrotransposons can be further divided into Ty1-*copia* and Ty3-*gypsy* classes, which differ from each other in their degree of sequence similarity (Xiong and Eickbush, 1990) and the order of encoded gene products (Kumar and Bennetzen, 1999). They are ubiquitous in plant genomes, constituting a significant proportion of the genome in many species (Bennetzen, 1996). For example, they represent at least 50% of nuclear DNA in maize (SanMiguel et al., 1996) and even in the smallest genomes, may still be present in high numbers (Arabidopsis-Genome-Initiative, 2002). Furthermore, retrotransposons may increase greatly in number over relatively short periods of evolutionary time (Pearce et al., 1996a). As such, variation in retrotransposon copy

number may account for a significant proportion of variation in DNA content as a whole (Kidwell, 2002). Indeed, variation in the number and variety of retrotransposons appears to be related to DNA content in a number of species (Kumar and Bennetzen, 1999). Furthermore, a study of intraspecific variation in the genomic copy number of BARE-1 family of retrotransposons in wild barley (*Hordeum spontaneum*) has suggested not only that such elements may be correlated with DNA content but that such variation may be adaptive (Kalendar et al., 2000).

A mechanism by which dispersed repetitive DNA may influence phenotype was proposed by the genome regulation theory (Meagher and Vassiliadis, 2005). Under this theory, dispersed repetitive elements may have both localised and genome wide effects through their influence on gene expression. Various mechanisms have been proposed for such an effect (see White et al., 1994; Zuckerlandl and Hennig, 1995; Vinogradov, 1998; Kato et al., 1999).

In *S. latifolia* repetitive DNA in particular is thought to be a significant factor in the relationship between flower size and DNA content within the species (Meagher and Costich, 1996). As such, a method for estimating the relative abundance of specific repetitive elements within the genus, and retrotransposons in particular, would be of great interest, and may provide valuable insights into the nature of the underlying genomic basis for this relationship. In particular, the role played by specific types of repetitive elements in influencing flower size in *Silene* is a question that remains to be addressed.

4.1.2. Estimating copy number

Traditionally, estimates of retrotransposon copy number have been made using either Southern blotting, or slot-blot methods (e.g. Pearce et al., 1996a; Kumar and Bennetzen, 1999). Southern blotting is considered most appropriate when copy numbers are low. Both methods are time consuming and the slot-blot method requires specialised equipment. An alternative technique, which would make it feasible and easy to rapidly survey large populations, would be extremely valuable.

4.1.3. PCR based estimation of copy number

There are a number of alternative methods by which variation in retrotransposon copy number could be detected. One such method is a quantitative fluorescence in situ hybridisation (FISH) approach but attempts to use this method in *Silene* have so far proved unreliable (Meagher & Yahr, personal communication). An alternative approach is to use a PCR based method.

Traditionally, PCR based methods for estimating the copy number of specific genomic elements have used either absolute or competitive techniques (Zimmermann and Mannhalter, 1996; Raeymaekers, 2000). Absolute methods compare the quantity of PCR product generated in a single PCR reaction to that of a standard curve of initial template versus product, whilst competitive methods compare the relative quantities of product derived from the target sequence with that derived from an internal standard (e.g. Honda et al., 2002). Whilst there are advantages to both of these techniques, there are also a number of drawbacks. Absolute methods such as quantitative PCR are expensive, require specialised equipment, and assume that all copies of the target sequence amplify with identical efficiencies. Competitive methods require a large amount of optimisation and rely on the assumption not only that the amplification efficiencies of the target sequences are similar, but also that the amplification efficiency of the internal standard and the target sequences are similar. Because amplification efficiency may depend on both the target and flanking sequence composition (Polz and Cavanaugh, 1998), these assumptions may not always be valid. Both of these methods assume that each copy of the sequence to be analysed is nearly identical, particularly at the primer binding sites. As such, these methods may have significant limitations for estimating copy numbers for relatively divergent or highly dispersed genomic elements such as retrotransposons. Nevertheless Meagher & Yahr (in press) have developed a quantitative PCR based method for estimating copy numbers of Ty1-*copia* class retrotransposons in *Silene*.

4.1.4. SSAP for estimating copy number

An alternative potential method for estimating copy numbers of retrotransposons, explored here, is a PCR based method known as site-specific amplification

polymorphism (SSAP) (Waugh et al., 1997). The technique is primarily used as a molecular marker system, and is similar in principal to the technique of amplified fragment length polymorphism (AFLP), as described in Vos *et al* (1995). The major difference between SSAP and AFLP is that rather than using two adapter primers only one is used, the other primer binding to a specific genomic sequence. The present study examines its potential as a method for estimating the copy number of Ty1-*copia* class retrotransposons based on the number of distinct PCR products generated in the SSAP PCR. Traditionally, SSAP marker systems have used only a subset of Ty1-*copia* retrotransposon sequences (those sharing similar LTR sequence). However, in order to examine overall copy number, the present study uses a conserved region of the Ty1-*copia* RNaseH gene. Vos et al (1995) noted that, in small genomes, AFLP band count was closely correlated with overall DNA content although the relationship was less strong in larger genomes, possibly due to the presence of repetitive DNA (Fay et al., 2005). Thus, it might reasonably be supposed that an SSAP based method could prove to be a useful technique for estimating the relative contribution of specific genomic elements to DNA content. The advantage of an SSAP based method over other PCR based approaches is that it does not rely on the assumption of equal amplification efficiency for each copy of the target sequence, requiring only that each PCR product is present in sufficient quantity to be detected. This is particularly relevant to analysis of plant retrotransposons which may contain significant sequence heterogeneity both within and immediately adjacent to the target sequence.

This method is not expected to provide an absolute estimate of copy number within an individual as not all copies of the target sequence will be expected to generate distinct PCR products, given that highly similar elements will share restriction enzyme cut sites. However, it may provide a reliable comparison of copy number between individuals.

4.1.5. Potential sources of error

Multilocus PCR based methods suffer from two classes of bias that can cause disparity between product and template ratios, namely, PCR selection and PCR drift (Wagner et al., 1994). The former can be defined as mechanisms which inherently

favour the amplification of certain templates due to properties of the genes, or their flanking sequences. These effects may include preferential denaturation due to low GC content in the target sequence or higher binding efficiency of GC rich variants of degenerate primers (Polz and Cavanaugh, 1998). Whilst there is no reason to suspect that variants of the target sequence being examined here will differ significantly in nucleotide base composition, the primer binding site (due to its short length) may vary significantly between variants of the target sequence. Bias introduced through PCR selection is expected to be reproducible between reactions using the same template and PCR conditions. PCR drift, on the other hand, is a stochastic effect whereby products amplified during the initial cycles of the PCR reaction, where amplification is predominantly from template DNA, can be disproportionately represented in the final pool of PCR products. Experimental studies have suggested that PCR drift is unlikely to be a significant factor in most multilocus PCR but that PCR selection may have significant effects (Wagner et al., 1994). The SSAP technique is, to some extent, resistant to the effects of PCR selection due to the fact that the technique detects only the presence or absence of distinct PCR products rather than their relative abundance. Nevertheless, if the effects of PCR selection cause the level of a particular product to fall below a certain threshold level, then it will not be detected during the analysis stage and will not be recorded. In order to investigate the extent to which PCR selection might affect the method examined here, sequence data incorporating the same RNaseH sequence motif, and generated using similar PCR conditions, are examined.

4.1.6. Aims

The aims of this study were to investigate whether an SSAP-like approach is an appropriate or useful method for detecting variation in copy number of Ty1-*copia* class retrotransposons within *Silene* and to compare the results obtained using this method with a quantitative PCR based method based on the same plants (Meagher & Yahr, in prep). In addition, in order to test whether Ty1-*copia* class retrotransposons contribute significantly to variation in genome size in the genus or whether they directly affect the flower size, the relationship between estimated copy number and DNA content and floral traits was examined. The potential effects of PCR selection upon the utility of the method were also tested.

4.2. Methods

4.2.1. Overview

An SSAP based approach was taken based on the protocol described in Pearce et al (1999). A conserved sequence motif from the Ty1-*copia* class retrotransposon was used as the site specific target. Genomic DNA was digested with restriction enzymes to generate fragments of varying lengths. Adapters were ligated to these fragments. Adapter and site specific primers were used in a PCR reaction. Thus, fragments containing the conserved sequence motif generated PCR fragments with a length corresponding to their proximity to the restriction enzyme cut site. The number of distinct fragment sizes was used to estimate genomic copy number of Ty1-*copia* retrotransposons.

4.2.2. Selection of plant material

Individuals were selected for the survey from four populations of *S. marizii*, which had previously been assayed for DNA content and flower size during the *S. marizii* phenotypic survey (described in Chapter 2) and two populations of *S. latifolia*, which had also been assayed for DNA content and floral traits. The estimates of DNA content amongst these plants allowed individuals to be selected for the study such that the survey included the extremes of DNA content variation within each population. Thus, although the survey may not be representative of the wider populations, it should have a greater power to detect any relationship between Ty1-*copia* copy number and phenotype, than with purely random selection. Sexual dimorphism in DNA content is a predominant feature characteristic of *S. latifolia*. In order to investigate the effect of sex upon the estimates of Ty1-*copia* abundance, approximately equal numbers of males and females from each population and species were selected for the survey.

4.2.3. DNA extraction

DNA extractions were made using a modified CTAB method (Lassner et al., 1989). Leaf material (200 mg) was collected in a 1.5 ml eppendorf tube. It was then frozen in liquid nitrogen and ground to a powder using a plastic pestle. 1 ml of 2xCTAB (0.1M Tris, 1.4M NaCl, 20mM EDTA, 2%CTAB, 1%PVP, 2% 2-B

mercaptoethanol) extraction buffer, preheated to 55°C, was added and mixed with the leaf material. The mixture was incubated at 55°C for 1 hour. The sample was left to cool for 10 minutes before being centrifuged at 13 000 rpm for 10 minutes. The supernatant was then removed. 600 µl of dichloromethane was added to the sample and subsequently mixed on a rotary shaker for 20 minutes. The sample was centrifuged at 13 000 rpm for 10 minutes and the supernatant removed. Again 600 µl of dichloromethane was added, the sample mixed on a rotary shaker for 20 minutes and centrifuged at 13 000 rpm for 10 minutes, before the supernatant was removed. 600µl of –20°C isopropenol was added to the sample and the tube inverted to precipitate the DNA. The sample was then centrifuged at 13 000 rpm for 10 minutes to form a pellet of crude DNA and the supernatant poured away. The sample was left to dry for 30 minutes, after which the pellet was dissolved in 500µl of TE (10mM Tris, 1mM EDTA). 3µl of RNase (10mg/ml) was added to the sample, which was then incubated at 37°C for 60 minutes. 50 µl of Sodium Acetate was added to the sample. 600 µl of –20°C, 96% ethanol was added to the sample and inverted to precipitate the DNA. The sample was then centrifuged at 13 000 rpm for 10 minutes to form a pellet of DNA. The supernatant was poured off and the sample left to dry for 30 minutes. The pellet was then dissolved in 200 µl of TE. DNA concentration was estimated by agarose gel electrophoresis with appropriate concentration standards.

4.2.4. Preparation of template DNA

PCR reactions used to generate the results were modified from those described in Pearce et al (1999), using RNaseH motif 2 as the conserved target sequence.

2.5µg of genomic DNA was completely digested with 1 unit of MseI in 30µl of reaction buffer (New England Biolabs). Reactions were incubated at 37°C for 1hr. Reactions were terminated by incubating at 80°C for 10 min to deactivate the restriction enzyme. Digested genomic DNA was purified using a qiaquick PCR purification column (Qiagen) according to the manufacturer's instructions.

MseI adapters were prepared by annealing two single stranded oligonucleotides: 25nmoles of MseI adapter 1 (5′-GACGATGGATCCTGAG-3′) and 25nmoles of

MseI adapter 2 (5′-TACTCAGGATCCAT-3′) were mixed in 250µl of annealing buffer (100mM Tris-HCl pH7.5, 1M NaCl, 10mM EDTA). Annealing was conducted by incubating the reaction at 65°C for 10min before leaving the sample to cool slowly to room temperature (1-2hrs). Template DNA was prepared by adding 2pmol of MseI adapters to 15µl of digested DNA in ligation buffer containing 1mM ATP and 0.1 units of T4 DNA ligase (New England Biolabs). The reactions were incubated at 4°C for 15hrs. Reactions were terminated by incubation at 70°C for 10 min. Template DNA was purified using a qiaquick PCR purification column in order to remove unligated adapter.

4.2.5. PCR of template DNA

50µg of template DNA was PCR amplified with 0.8µg of dye labelled RNaseH motif 2 primer, labelled with D4 dye (Beckman Coulter) (5′-GCNGAYATNYTNACNAA-3′) and 0.15µg MseI adapter primer (5′-GATGGATCCTGAGTAAGATG-3′) in 50µl 0.2mM deoxynucleotide triphosphates (dNTPs), 1x PCR reaction buffer, 0.25 units of hotstar Taq DNA polymerase (Qiagen). Preparation of PCR reactions was conducted in a PCR hood, which had been sterilised by UV irradiation. The MseI adapter primer included 4 selective bases to cut down on the number of unique PCR products, which may otherwise have numbered in the thousands. Thermal cycling was conducted using an Applied Biosystems Geneamp PCR machine, and consisted of 15min at 95°C prior to 30 cycles of: 1min at 94°C; 1min 30s at 45°C; and 1min 30s at 72°C.

4.2.6. PCR fragment analysis

PCR fragments were analysed using a CEQ 8000 DNA sequencer (Beckman Coulter) in order to separate fragments of different lengths. 0.5 µl of PCR product was added to 40µl of formamide containing a 400bp dye labelled size standard, and the mixture transferred to a single well on a 96 well PCR plate. The sample was then processed on the sequencer.

The raw data were analysed using the fragment analysis module of the CEQ system software. Signal peaks were filtered according to a number of criteria: only signal

peaks with a maximum intensity that was 10% or more of that of the maximum signal peak were included. Similarly, only peaks with a slope parameter of greater than 10 were identified (the slope parameter is a measure of the rate of increase in signal intensity for a given peak, which is dependent on the height and width of the peak, although its absolute value is dependent on the baseline noise). Fragment sizes for each peak were estimated based on an internal size standard. Peaks were assigned to 'bins' that were within half a base pair of the nominal size. The size interval between adjacent bins was one base pair. The number of occupied bins for each sample was recorded. Ty1-*copia* retrotransposon copy number was estimated by multiplying the observed peak number by 256 (the expected reduction in PCR fragments due to the incorporation of 4 selective bases in the MseI adapter primer).

4.2.7. SSAP peak numbers

The relationships between various traits and SSAP peak numbers were examined using Pearson correlation coefficients. SSAP peak number was compared with overall DNA content (determined by flow-cytometry), three floral traits: mean limb length, mean claw length and mean calyx diameter, and a quantitative PCR based method estimate of Ty1-*copia* copy numbers. qPCR based estimates (Meagher & Yahr, in prep) deviated significantly from a normal distribution, and a logarithmic transformation was applied to these data prior to analysis. The effect of species, population and sex upon SSAP peak number was determined using a partially nested ANOVA, which examined the effects of species, population (nested within species) and sex.

4.2.8. Base composition analysis

One potential confounding effect on multilocus PCR is that of PCR selection, particularly when (as in the method employed here) degenerate primers are used. In order to investigate the extent to which PCR selection may affect the results of the SSAP analysis, the base composition of RNaseH motif 2 for two different types of sequence were analysed. These were:

1. Sequence data which used RNaseH motif 2 as a primer binding site.

2. Sequence data which incorporated RNaseH motif 2, but used motif 1 as a primer binding site.

These sequences were obtained from a number of sources. Sequences from *S. marizii* and *S. latifolia* were produced during the attempts to develop molecular markers from retrotransposons (described in Appendix B). This amounted to 26 sequences, all of which used motif 2 as the conserved primer binding site. Also included were previously published RNaseH sequence data, from a variety of species. This provided an additional 37 sequences (23 of which used motif 2 as primer binding site). These sequences were all taken from two previous publications (Pearce et al., 1996a; Pearce, 2007) and were retrieved from the EMBL nucleotide sequence database (Kulikova et al., 2007). The retrieved sequences derived from four species: *Pisum savitum*, *Vicia faba*, *Picea abies* and *Glycine max*. All of these sequences are summarised in Appendix A.

4.2.9. Sequence analysis

In order to investigate the extent to which PCR selection affects the results of SSAP analysis, the nucleotide base composition of RNaseH motif 2 was determined for each sequence. The base composition was pooled for guanine and cytosine as GC content is generally considered to be the most significant factor in PCR selection (Polz and Cavanaugh, 1998). Sequences were divided into three groups for the purposes of analysis based on the source of the sequence (Pearce, et al publications or *Silene*) and the location of the primer site (motif 1 or motif 2).

All possible primer sequences were determined, based on the degenerate primer sequence: 5′-GCNGAYATNYTNACNAA-3′. This gave 1024 equally probable sequences. The nucleotide base composition was determined for each of these theoretically possible primers. Although it is not possible to know the true frequency distribution of GC content for the population of primers, it was assumed that the mean was identical to the mean of the theoretical range (i.e. under the assumption that each degenerate base is equally likely). This assumption may not be entirely valid as the primer manufacturers (Invitrogen) do not guarantee equal

mixing of bases at degenerate sites. However, it is reasonable to suppose that base composition at each of the six degenerate sites is not biased towards any particular base pair.

4.2.10. Statistical analysis

If PCR selection is a significant effect in the SSAP method, then one would expect that sequences incorporating RNaseH motif 2 would have a different bases composition to those that use this motif as a primer site. An ANOVA was used to test whether the base composition of the RNaseH motif 2 sequence was affected by sequence type. Comparisons between the mean of each sequence type was made using Tukey's method with a family error rate of 5%.

4.3. Results

4.3.1. Variation in SSAP peak number

Ten individuals were assayed from two populations of *S. latifolia*; six females and four males. Eleven individuals from four populations of *S. marizii* were also included in the study; five females and six males. Representative examples of SSAP output are shown in Figure 4.1. SSAP peak numbers ranged from 4 to 35; if it is assumed that every potential target generated a distinct and detectable peak, this would correspond to copy numbers of between 1024 and 8960, when the selectivity of the adapter primer is accounted for. Variation in SSAP peak numbers was considerable both within species and within populations (Table 4.1).

Table 4.1: Means and standard deviations for SSAP peak number by a: population and b: sex.

a.

Species	Population	N	Mean	StDev
S. latifolia	1	5	14.8	10.01
	7	8	21.5	11.17
S. marizii	8	2	11.5	4.95
	15	1	26.0	-
	16	2	17.0	0.00
	17	3	21.0	8.72
	18	2	17.5	13.44

b.

Species	Sex	N	Mean	StDev
S. latifolia	Female	6	22.3	12.24
	Male	6	16.2	10.32
S. marizii	Female	5	20.0	9.11
	Male	5	16.2	6.46

4.3.2. Effect of species, sex and population

There was no evidence for an effect of species on SSAP peak number. Neither was there evidence for an effect of population, nested within species, or between sexes (Table 4.2).

Table 4.2: Partial hierarchical analysis of variance (ANOVA) results testing for differences in SSAP peak number among species, populations, and sexes.

Source	Degrees of freedom	Sequence sum of squares	Adjusted sum of squares	Adjusted mean square	*F*-ratio	*p*
Species	1	111.1	98.6	98.6	0.91	0.353
Population (Species)	4	417.5	440.3	110.1	1.01	0.425
Sex	1	128.9	128.9	128.9	1.18	0.290
Error	20	2177.6	2177.6	108.9		
Total	26	2835.2				

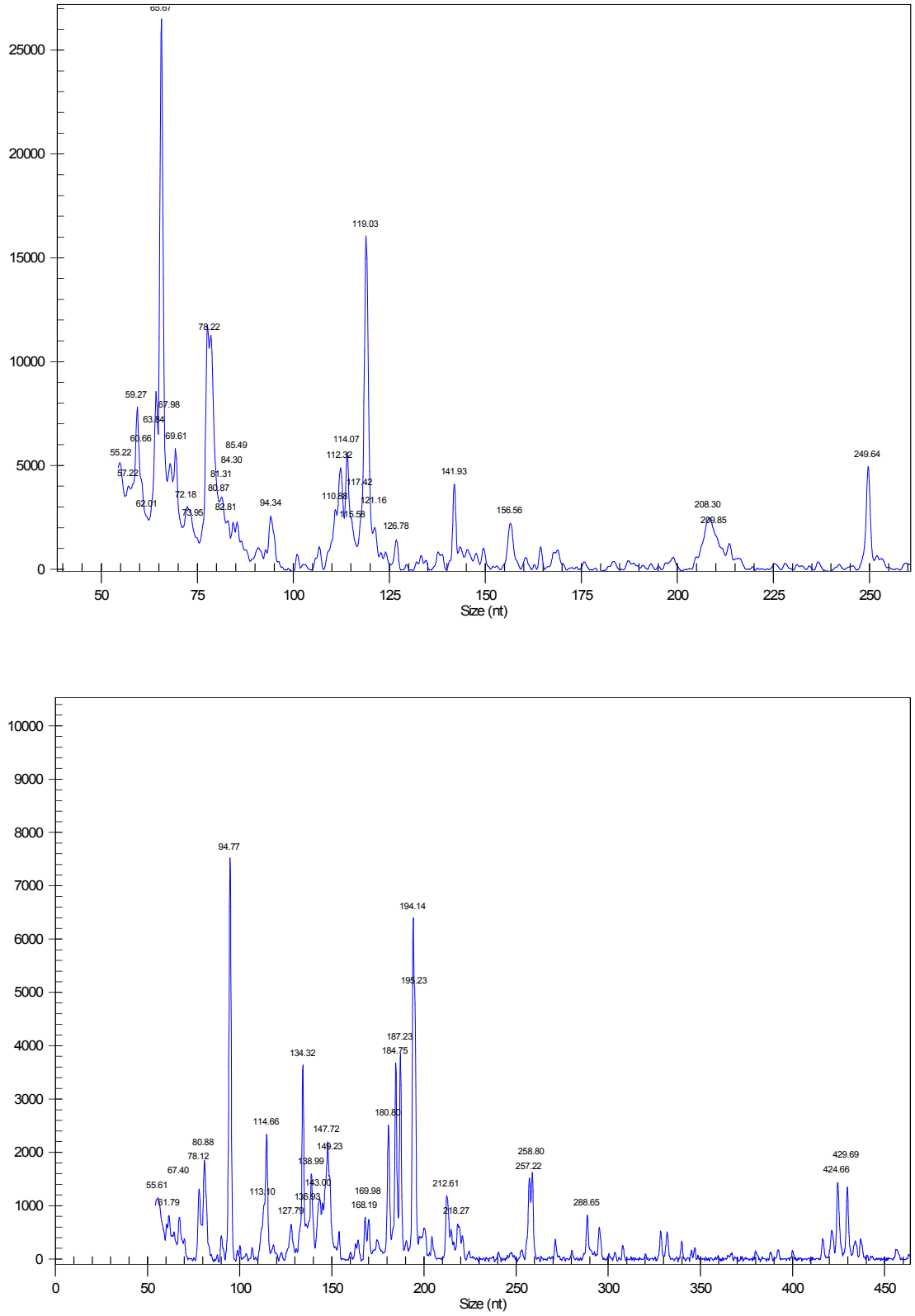


Figure 4.1: Example traces from an SSAP sample run. Peaks show SSAP fragments; detected fragments are indicated by their size. The upper trace is from individual 1-10-5 and the lower from PT-01-01. The x axis represents fragment size (in base pairs) and the y axis shows signal intensity.

4.3.3. Correlations

There was a significant negative correlation between SSAP peak number and DNA content. Correlation coefficients, for this relationship, were consistent between both species and sexes although the correlation was only significant when the data were pooled (Table 4.3, Figure 4.2). There was no significant correlation between the qPCR based, and SSAP based estimate of Ty1-*copia* copy number, although there was a trend towards positive correlations in all cases (Table 4.3, Figure 4.2). None of the three flower size measurements showed any evidence for a significant correlation with SSAP peak number (Table 4.3, Figure 4.2).

Table 4.3: Correlations by species and sex between SSAP peak number and phenotype.

Trait		***S. latifolia***		***S. marizii***		**Overall**	
		r	***p***	***r***	***p***	***r***	***p***
DNA Content	Male					-0.337	0.342
	Female					-0.526	0.097
	Overall	-0.467	0.148	-0.326	0.358	-0.445	0.043
Ln Copy Number (qPCR)	Male					0.390	0.299
	Female					0.167	0.668
	Overall	0.253	0.452	0.495	0.212	0.354	0.137
Limb length	Male					0.317	0.372
	Female					-0.123	0.736
	Overall	0.264	0.461	-0.138	0.703	0.087	0.716
Claw length	Male					0.097	0.790
	Female					0.311	0.381
	Overall	0.063	0.864	-0.159	0.661	0.196	0.409
Calyx Diameter	Male					0.216	0.550
	Female					0.175	0.628
	Overall	0.263	0.462	0.136	0.708	0.304	0.192

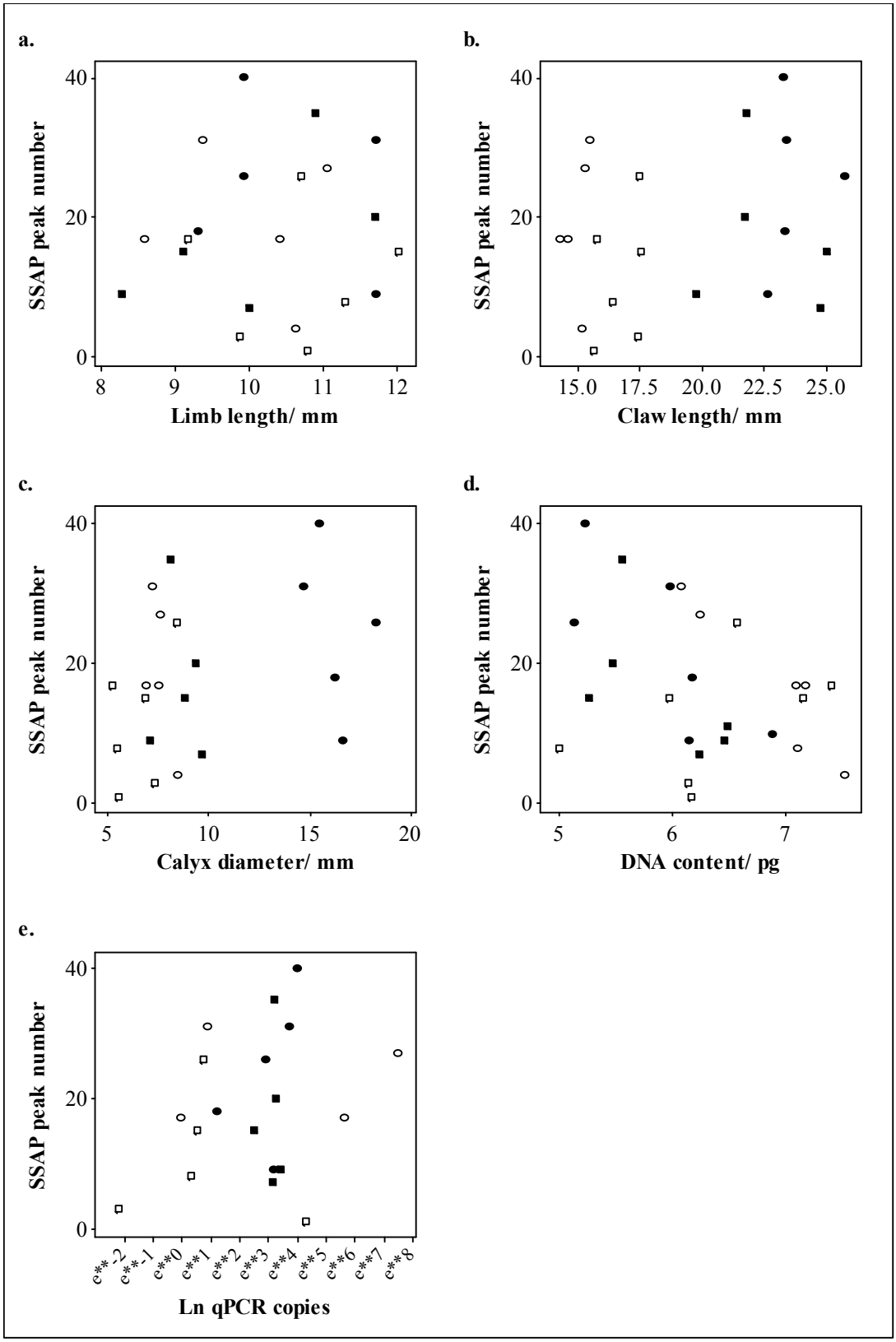


Figure 4.2: Relationships between SSAP peak number and individual means for the three floral traits and DNA content. Filled symbols represent *S. latifolia*, and open symbols *S. marizii*. Squares represent males, and circles females.

4.3.4. Base composition analysis

RNaseH motif 2 sequences from PCRs that used RNaseH motif 2 as a primer binding site contained a greater proportion of GC bases than sequences that used motif 1 as a primer binding site. This was true for both for *Silene* sequences and for sequences from other species (retrieved from the EMBL nucleotide sequence database). The sequences retrieved from EMBL had a significantly lower GC content than the sequences from *Silene*, regardless of the primer site used (Table 4.5, Table 4.6).

Table 4.4: The mean and standard deviation of GC content in the three samples of RNaseH motif 2 sequence.

Source	Primer site	N	Mean	St Dev
Looseley	Motif 2	26	8.3	0.96
Pearce	Motif 2	23	6.9	1.40
Pearce	Motif 1	14	5.3	0.83

Table 4.5: ANOVA table examining the effects of sequence type on GC content.

Source	Degrees of freedom	Sum of squares	Mean square	*F*-ratio	*p*
Sequence Type	2	82.42	41.21	37.58	0.000
Error	60	65.80	1.10		
Total	62	148.22			

Table 4.6: Comparison of the mean number of GC bases in RNaseH motif 2 of the three sequence types using Tukey's method. Table elements show the differences between the mean of the corresponding row and column heading. Statistically significant differences, at the individual confidence level of 0.98, are marked with an asterisk.

	Pearce motif 2	Looseley motif 2
Looseley motif 2	1.36 *	
Pearce motif 1	1.63 *	3.00 *

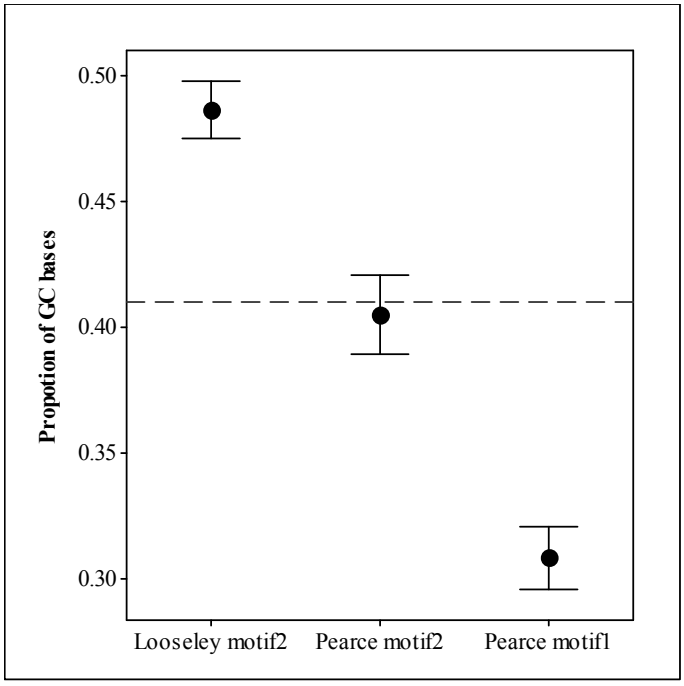


Figure 4.3: Means and standard errors of the GC content of RNaseH motif 2 for the 3 different sequence types. The theoretical mean of the primer population (0.41) is indicated by the dashed line.

4.4. Discussion

4.4.1. Variation in copy number

The results of the SSAP survey suggest that there is a great deal of variability in the copy number of Ty1-*copia* class retrotransposons within the individuals surveyed. This is in general agreement with the results obtained using the quantitative PCR based approach. Nevertheless, whilst it is conceivable that considerable differences in retrotransposon copy number might exist between species or between sexes, it is difficult to envisage how such high variation in copy number could be maintained within a single population, given interbreeding and realistic levels of transpositional activity (Flavell et al., 1997). As such, it seems likely that a significant proportion of this variation may be due to error in the estimates rather than true variation in copy number. This is discussed in more detail below.

4.4.2. Species and sex effects

The lack of sexual dimorphism in SSAP peak number is consistent with the observation by Matsunaga et al (2002) that Ty1-*copia* class retrotransposons do not accumulate preferentially on the Y-chromosome. The absence of a species effect on the SSAP peak number suggests that variation in Ty1-*copia* class retrotransposons does not account for a significant proportion of the variation in DNA content between these two species. Nevertheless, it is possible that the likely error in the estimates of copy number would mask any such effects that might exist.

4.4.3. Relationship between SSAP peak number and qPCR estimates

There is only a very general trend towards a positive relationship between the two methods for estimating copy numbers. There are a number of reasons why this might be so. Firstly, the SSAP method detects distinct copies only when there is a polymorphism between the two copies of the target sequence at the restriction enzyme binding site, whilst the qPCR method should be able to distinguish between several very similar elements. As such, estimates using the SSAP method may better reflect the number of retrotransposon subgroups rather than the number of individual copies. Another significant difference between the two methods is that the qPCR based method uses two retrotransposon specific primers, whilst the SSAP method uses only one. This would render the SSAP based approach more efficient at detecting inactive retrotransposon copies, being less dependent on binding site specificity. In addition, the SSAP technique (unlike the qPCR based approach) uses a degenerate PCR primer based on the amino acid sequence of RNaseH motif 2. As such, it is less susceptible to bias in the frequency of underlying sequence variants.

4.4.4. Correlation with DNA content

The significant negative correlation between DNA content and SSAP count appears counterintuitive; one would naturally expect individuals with more retrotransposon copies to have larger genomes, an expectation that is supported by a number of studies (e.g. Kumar, 1996). However, it is worth noting that the class of retrotransposon being examined here represent only a fraction of overall DNA

content, and potentially only a small fraction of the repetitive DNA. In addition to this, it is unclear how effective the method employed here is at detecting inactive retrotransposon copies; much of the effect of retrotransposon activity on DNA content may be due to historical activity. Indeed, assuming that there is an upper limit on the host genomes tolerance for retrotransposon insertions then suppression of retrotransposon activity and therefore the extent of active (or recently active copies) may be greater in large genomes.

An alternative, methodological, explanation might be that individuals with many potential target sequences (e.g. with large genomes) might show lower average peak heights, which could, potentially, disqualify peaks that would otherwise be detected by the analysis. However, given the selectivity of the adapter primer, and the generally low numbers of peaks, it seems unlikely that availability of substrate will be a limiting factor in these reactions.

4.4.5. Potential sources of error and bias

As noted above, it seems likely that there may be significant error associated with the SSAP method for determining copy number. The SSAP method requires that each target sequence generates enough PCR products to produce a detectable peak. PCR selection is one mechanism which might prevent this in some targets. Indeed, the strong bias towards high GC content in sequences using the RNaseH motif 2 primer site suggests that this may be the case. Nevertheless, it remains possible that the GC bias, in sequences that use RNaseH motif2 as a primer binding site, is simply a reflection of an underlying sequence bias in the RNaseH gene of the Ty1-*copia* retrotransposons being examined. Although it is not possible to test whether this is the case from the available data in *Silene*, RNaseH sequence data from other plant species show that there is no GC bias in the underlying sequence, but that a significant bias is present when motif 2 is used as a primer site. These results suggest that PCR selection is a major influence on the composition of products from this PCR.

The strong bias towards GC rich target sequence may cause a downward bias in the estimates of copy number obtained using this method, but it would be expected to

act consistently between samples, and thus is unlikely to explain the high variability in SSAP peak number. As such, the high variability seen in these results is more likely to be caused by variation in some other factor. Variation in template quality or quantity are alternative explanations for this effect.

4.4.6. Summary

These results suggest that the high levels of variation in SSAP peak number represent methodological error rather than real variation in retrotransposon copy number. This is likely to render the method unreliable as a means of comparing copy numbers amongst large survey populations. The finding of strong bias towards GC rich variants of the target sequence suggests that estimates of copy number using this method may be biased downwards. As such, the role of these elements in genome evolution within *Silene*, or their effects on floral variation is a question that remains to be addressed. The finding that PCR selection has a strong influence in the PCR reactions described here is a general result that is applicable to any study using degenerate primers, particularly when sequences derived from such reactions are assumed to represent random samples from the underlying population of sequences. This is discussed with regard to the phylogenetic analysis of Ty1-*copia* retrotransposons described in the next chapter.

Chapter 5 Evolutionary dynamics of Ty1-*copia* class retrotransposons in plant genomes

5.1. Introduction

5.1.1. Evolutionary dynamics of Ty1-*copia* retrotransposons

Ty1-*copia* class retrotransposons are ubiquitous in plant genomes and may represent a significant fraction of the genome in many species (Pearce et al., 1996a; SanMiguel et al., 1996; Kumar and Bennetzen, 1999), with large scale increases in copy number arising over relatively short periods of time e.g. in maize (SanMiguel et al., 1996; SanMiguel et al., 1998) or *Vicia* species (Pearce et al., 1996a). As such, variation in retrotransposon copy number may be a significant determinant of overall variation in DNA content (Kidwell, 2002). Therefore, the evolutionary dynamics of these elements is of considerable relevance to the study of genome size evolution. Theoretical studies predict that levels of sequence heterogeneity between retrotransposon copies should be related to genomic copy number because each element has a given probability of mutating (Charlesworth, 1986). This suggestion has been validated amongst *Vicia* species which vary considerably in genomic copy number (Pearce et al., 1996a), although the same study found no relationship between sequence heterogeneity and DNA content. As such, comparisons of sequence heterogeneity may provide important clues to relative abundance of these elements between species.

A number of previous studies have examined Ty1-*copia* class retrotransposons in a wide variety of plant genomes. These have generally studied sequence data from the reverse transcriptase gene (e.g. Flavell et al., 1992a; Pearce et al., 1996a; Matsuoka and Tsunewaki, 1999; Stuart-Rogers and Flavell, 2001; Matsunaga et al., 2002; Navarro-Quezada and Schoen, 2002) and have, without exception, found that these retrotransposons are subject to strong purifying selection, but are present in large, highly heterogeneous populations. This heterogeneity has been accounted for by generally low levels of transpositional activity, coupled with a high tolerance for retrotransposon accumulation in plant genomes (Flavell et al., 1997).

5.1.2. Relationship between element and species phylogeny

Ty1-*copia* type retrotransposons lack the envelope (env) gene that is required for cell to cell transmission (Kumar and Bennetzen, 1999). This means that the evolution of families of retrotransposons is expected to match that of the host genome. Indeed, there have been no convincing examples documenting horizontal transmission of such elements in plants (Flavell et al., 1997; Kumar and Bennetzen, 1999) and, as such, genetic distances between the most similar sequences from two species are expected to be proportional to the evolutionary distance between the species. This has allowed the phylogenetic analysis of Ty1-*copia* retrotransposon sequence to elucidate the evolutionary relationship between species (e.g. Hirochika and Hirochika, 1993; Matsuoka and Tsunewaki, 1996; Pearce et al., 1996b; Matsunaga et al., 2002). As such the phylogenetic analysis of these sequences may shed light on the level of genetic divergence between *S. latifolia* and *S. marizii* and the likely extent to which hybridisation and, particularly, introgression exists between these species.

5.1.3. Aims

In this study, sequence data from the Ty1-*copia* RNaseH gene were subjected to phylogenetic analysis to examine patterns of sequence variation and evolution within this gene. The study examined sequences obtained both in *S. latifolia* and *S. marizii* and a number of previously published sequences from a variety of other species. A particular aim of the study was to quantify the level of heterogeneity

among these sequences, and to relate these to the action of selection within the sequence. Levels of heterogeneity are also compared with DNA content and with previous estimates of Ty1-*copia* copy number within each species. Finally, the results of the phylogenetic analysis were interpreted with regard to the likely extent of genetic isolation between *S. marizii* and *S. latifolia*.

5.2. Methods

5.2.1. Sequences

Partial sequences of the Ty1-*copia* RNaseH gene were generated and sequenced from *S. marizii* and *S. latifolia* using a method based on that described in Pearce (1999) (see Appendix B for more details). Also included in the analysis were a number of partial RNaseH sequences, from other species, that were retrieved from the EMBL nucleotide sequence database (Kulikova et al., 2007) from two previous publications (Pearce et al., 1999; Pearce, 2007). The origin of all these sequences is listed in appendix A.

5.2.2. Sequence alignment

It was assumed that alignments based on amino acid sequence were the most appropriate way to treat these data. As such, nucleotide sequences were translated into amino acid sequence using Genedoc (Nicholas and Nicholas, 1997). Sequence data beyond the termination of the open reading frame were discarded as an amino acid alignment on these data would not be valid. In addition, sequences which had very low levels of conservation relative to the consensus sequence were discarded; sequences were only retained if they contained at least 5 conserved bases (bases present in more than 80% of sequences) in the first 65 positions. Finally, sequences which contained stop codons within 62 bases of the first base position were discarded. Both of these steps were taken because such sequences were considered likely to represent inactive retrotransposon copies. Such sequences were not included because, again, it was not considered appropriate to include these sequences in an amino acid alignment. Following these measures, the resulting data

set consisted of 43 sequences from 5 species (one gymnosperm and four dicots). The species represented were*: S. marizii* (8 sequences), *S. latifolia* (14 sequences), *Vicia faba* (broad bean) (6 sequences), *Pisum savitum* (pea) (9 sequences) and *Picea abies* (Norway spruce) (6 sequences). Amino acid sequences were aligned using ClustalX software (Thomson et al., 1997) employing the default alignment parameters. Finally, the aligned amino acid sequences were converted back to nucleotide sequence (using the original codons) for phylogenetic analysis.

5.2.3. Phylogenetic analysis

In order to reconstruct the phylogeny under a realistic evolutionary model, a Bayesian phylogenetic analysis was conducted using the programme MRBAYES (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). Bayesian inference of phylogeny is based on the posterior probability of the most probable tree, given the aligned sequence data. Posterior probabilities cannot be determined analytically, so a Metropolis-coupled Markov Chain Monte Carlo (MCMCMC) algorithm is used. The Markov Chain Monte Carlo method is an iterative procedure, in which the acceptance of a new state is dependent on the likelihood ratio of the old and new states. By sampling from the trees produced by such a procedure, the posterior probabilities of particular phylogenetic partitions can be approximated. The Metropolis coupled variant of this procedure uses multiple iterative chains, a number of which are heated (such that state swaps are more likely). By allowing state swaps between the heated and unheated chains (from which samples are taken), this variant can more easily explore the likelihood surface of a given dataset than the traditional Markov Chain Monte Carlo method.

The evolutionary model used for likelihood calculations was the default model applied by MRBAYES except that the substitution rates for transversions and transitions were included as separate parameters. In addition, substitution rates across sites were considered to be variable; the rate at a given site being drawn from a gamma distribution. This model was considered to be a reasonable compromise between maximising the realism of the evolutionary model, whilst minimising the number of parameters.

The MCMCMC algorithm was run for $2x10^6$ iterations. The analysis conducted two independent runs, each incorporating 1 cold chain and 3 heated chains. Trees were sampled from the cold chains every 4000 iterations. This frequency was used to ensure that sampled trees were (as far as possible) independent. This independence of trees was tested by performing an autocorrelation analysis on the log-likelihoods of the sampled trees. The sampling frequency was accepted if the probability of the autocorrelation, with a lag of 1, was less than 0.05. Log-likelihood scores for each of the sampled trees were recorded. Trees that were sampled from iterations prior to the convergence of the log-likelihoods on a stable value were not used for the production of the consensus tree.

As well as a Bayesian analysis, the aligned sequence data were also analysed using a neighbour-joining procedure as implemented in the ClustalX software. This allowed for a comparison of the phylogenetic trees produced using two entirely separate methods.

5.2.4. Analysis of selection

Fundamental to the interpretation of the phylogenetic analysis is the strength and nature of the selective forces acting on the sequences examined in the study. In order to investigate the nature of selective forces acting on these sequences, pairwise estimates of the rate of synonymous substitutions (K_s) and non-synonymous substitutions (K_a) were made. The ratio of these estimates (K_a/K_s) give an indication of the strength and direction of selective forces, with values greater than one indicating directional selection and values less than one indicating purifying selection (Kimura, 1983). Sequences were paired according to their postulated relationships as determined by the phylogenetic analysis such that each sequence were paired with its closest relative. These pairings were made to minimise the divergence between sequences in each pairwise comparison which might otherwise underestimate the rates of divergence due to multiple substitutions at a single site. Furthermore, it seems reasonable to assume that selection is likely to be more similar between similar sequences. However, this approach meant that a number of sequences were represented more that once in the analysis.

Rates of synonymous divergence and non-synonymous divergence were estimated by a maximum likelihood procedure as implemented in the codeml program of the PAML software package (Yang, 1997), under the F3x4 model of codon substitution. A maximum likelihood approach was used because they are generally thought to suffer fewer biases than approximate methods (e.g. Yang and Nielsen, 2000), although this can be reversed when very short sequences are examined (Yang and Nielsen, 2000). The likelihood of the data was recorded under a null model of $K_a/K_s=1$ and under an alternative model where K_a and K_s were allowed to vary independently. The statistical significance of the K_a/K_s ratio estimated by the alternative model was tested using a likelihood ratio test. The test statistic was calculated as: $LR = 2(\ln M_0 - \ln M_1)$, where M_0 is the likelihood under the null model and M_1 is the likelihood under the alternative model. The test statistic was compared to a χ^2 distribution with one degree of freedom to obtain the *p*-value.

5.2.5. Sequence heterogeneity

Heterogeneity of sequences, within each species, was quantified by calculating the mean within species branch lengths to give the expected divergence between sequences within a species. In addition to this, the genetic distance between the two most similar sequences from each species comparison was calculated in order to examine evolutionary relationships between the species.

Theoretical studies predict that retrotransposon sequence heterogeneity should be related to copy number (Charlesworth, 1986). In order to compare levels of sequence heterogeneity with estimates of DNA content and copy numbers of Ty1-*copia* retrotransposons, estimates of haploid nuclear DNA content were retrieved from the C-value database maintained by the Royal Botanic Gardens Kew (Bennett and Leitch, 2004). Estimates of copy number were retrieved from published studies and, within *Silene*, were taken from within species means from the SSAP based method (described in Chapter 4).

5.3. Results

5.3.1. Phylogenetic analysis

The consensus tree generated by MRBAYES is shown in Figure 5.1. In general, the cladistic assignments of this tree are very similar to those generated by the neighbour joining method (Figure 5.2), but the branch lengths are substantially greater. This is almost certainly because the neighbour joining method did not correct for multiple substitutions at a given site. This underlines the importance of incorporating multiple substitutions for analyses of highly divergent sequences. The apparently higher resolution of the neighbour joining tree for deeper divisions is likely to be a consequence of the method employed rather than reflecting any real ability to resolve these relationships. This is demonstrated by the very low levels of bootstrap support for these partitions. In a number of cases, sequences from very distantly related species appear to group together, in particular sequences from the single gymnosperm (*P. abies*) and sequences from *P. savitum* and *V. faba*. However, there are, without exception, very low levels of support for these partitions and it seems unlikely that these partitions represent real relationships.

Ty1-*copia* class retrotransposons were present as a highly heterogeneous population in the species examined, with a number of defined and well supported subgroups (Figure 5.1). Comparisons of similar sequences within and between species showed that sequences within species were more similar than when interspecific comparisons were made (Table 5.1). To some extent, differences in these comparisons reflect the inequalities in sampling between species. Nevertheless, the genetic distance between the most similar *S. latifolia*/ *S. marizii* comparison was substantially greater than the comparisons from within the two species, indicating considerable isolation between retrotransposon populations in the two species.

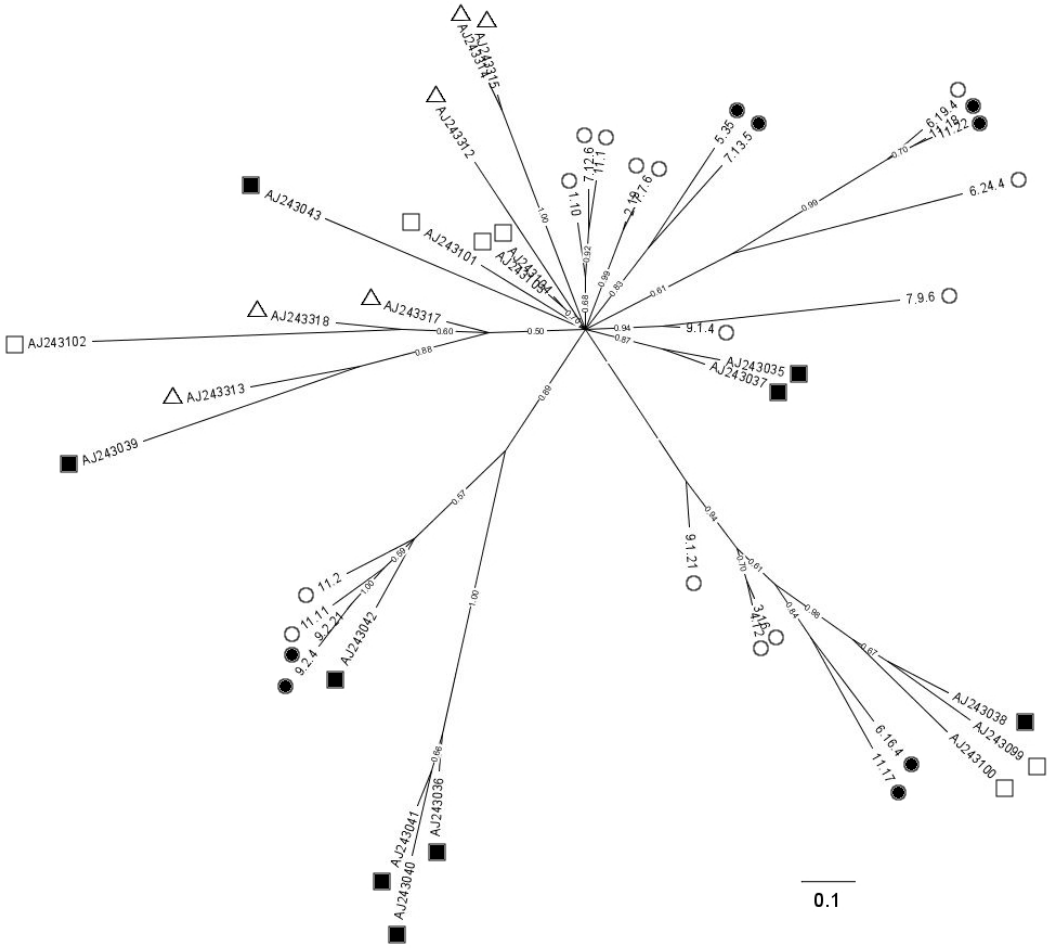


Figure 5.1: MRBAYES consensus tree of evolutionary relationships between the sequences included in the study. Numbers prior to each partition represent the posterior probability of the partition. The horizontal line corresponds to a branch length of 0.1 substitutions per site. Filled squares represent sequences from *P. savitum*; open squares: *V. faba*; open circles: *S. latifolia*; filled circles: *S. marizii* and open triangles: *P. abies*.

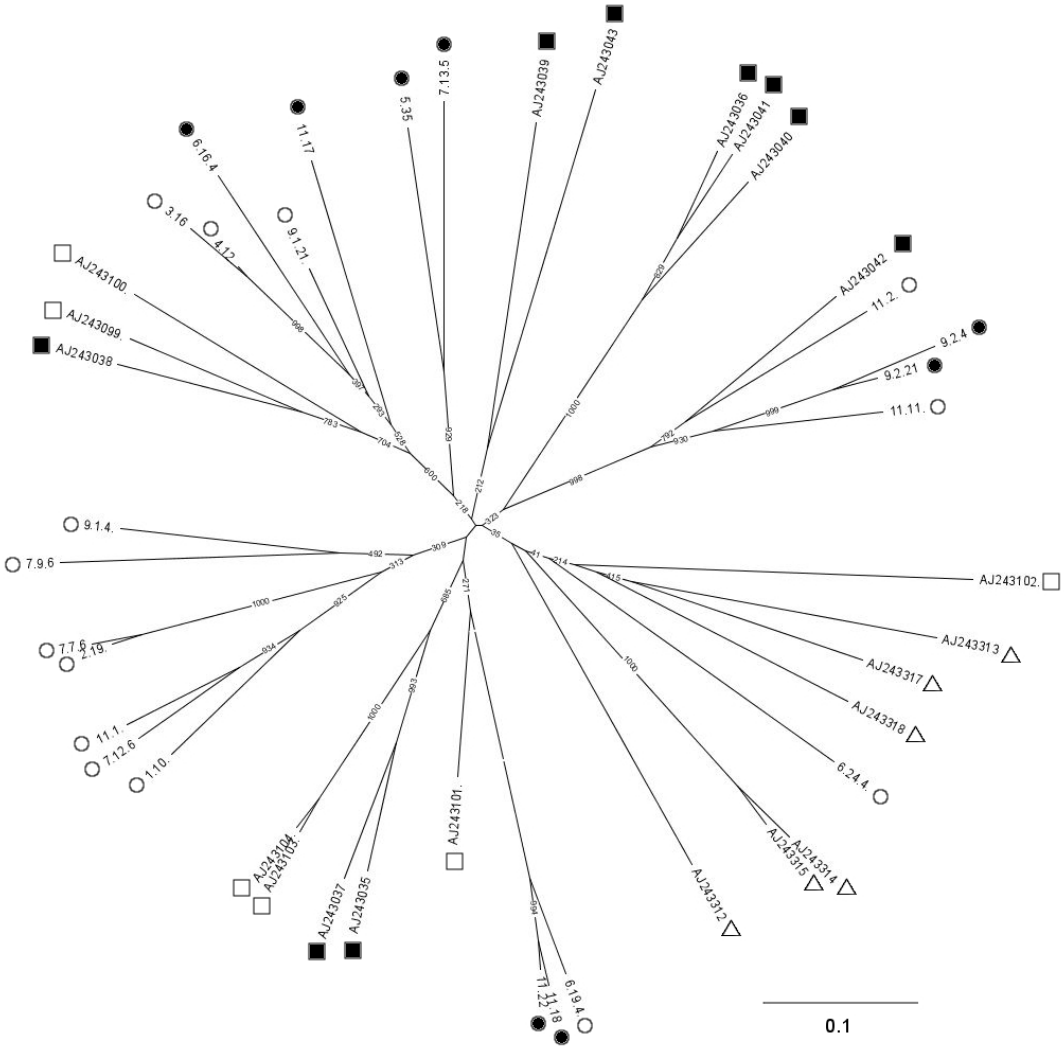


Figure 5.2: Clustal X neighbour joining tree of evolutionary relationships between the sequences included in the study. Numbers prior to each partition represent the number of bootstrap replicates (out of 1000) that support the partition. The horizontal line corresponds to a branch length of 0.1 substitutions per site. Filled squares represent sequences from *P. savitum*; open squares: *V. faba*; open circles: *S. latifolia*; filled circles: *S. marizii* and open triangles: *P. abies*.

Table 5.1: Genetic distance between the most similar sequences from each species comparison, defined as the number of expected changes per site.

	S. latifolia	*S. marizii*	*P. savitum*	*V. faba*	*P. abies*
S. latifolia	0.073	0.152	0.265	0.274	0.449
S. marizii		0.053	0.469	0.463	0.638
P. savitum			0.172	0.158	0.332
V. faba				0.051	0.342
P. abies					0.078

5.3.2. Analysis of selection

For almost all of the pairwise comparisons, the K_a/K_s ratio was less than 1, indicating the action of purifying selection on the sequences (Table 5.2). In a number of cases the ratio was greater than 1, but this was generally associated with a low rate of total substitution caused by short branch lengths between sequences, which are unlikely to provide a reliable estimate of K_a/K_s. Indeed this is reflected by the low *p*-values associated with such comparisons.

Table 5.2: Results of pairwise comparisons of synonymous and non-synonymous rates of substitution between sequences. S is the number of synonymous sites and N is the number of non-synonymous sites. The estimate of the ratio K_a/K_s is shown along with the probability that K_a/K_s is different to 1.

Sequence 1	Sequence 2	S	N	K_a/K_S	LR	*p*
AJ243035	AJ243037	19.7	52.3	0.110	9.151	0.003
AJ243036	AJ243041	14.6	48.4	1.332	0.063	0.802
AJ243040	AJ243041	14.5	48.5	0.251	2.803	0.094
AJ243042	11.2	19.1	52.9	0.112	12.685	<0.001
9.2.4	9.2.21	13.6	58.4	0.384	1.269	0.260
11.11	11.2	20.3	51.7	0.047	21.295	<.0001
AJ2430038	AJ243099	17.6	54.4	0.002	29.672	<.0001
AJ243100	AJ243099	15.9	56.1	0.006	22.717	<.0001
6.16.4	11.17	21.4	50.6	0.096	13.922	<0.001
3.16	4.12	14.5	57.5	0.048	7.649	0.006
9.1.21	3.16	13	50	0.365	1.228	0.268
AJ243039	AJ243313	15.2	56.8	0.007	17.091	<.0001
AJ243318	AJ243102	16.8	55.2	0.485	0.854	0.356
AJ243317	AJ243318	15.6	56.4	0.079	15.612	<.0001
AJ243043	AJ243103	17.1	54.9	0.007	17.295	<.0001
AJ243101	AJ243104	6	27	0.023	10.512	0.001
AJ243312	AJ243103	14.8	57.2	0.045	14.112	0.000
AJ243314	AJ243315	18.9	53.1	2.144	0.528	0.467
1.1	7.12.6	17.5	54.5	0.196	5.557	0.018
7.12.6	11.1	20.4	51.6	0.198	5.822	0.016
2.19	7.76	20.7	51.3	0.167	4.265	0.039
5.35	7.13.5	19.5	52.5	0.164	9.617	0.002
6.19.4	11.18	19.6	52.4	1.079	0.004	0.949
11.18	11.22	21.4	50.6	0.199	1.761	0.185
6.24.4	11.18	15.8	56.2	0.007	4.379	0.036
9.1.4	7.9.6	18.9	53.1	0.161	7.867	0.005

5.3.3. Sequence heterogeneity, DNA content and copy number

Overall, there was no significant relationship between estimates of heterogeneity and either copy number or DNA content (Table 5.3, Figure 5.3), although when genomes of similar size considered, there is a trend towards a positive relationship with DNA content.

Table 5.3: Estimates of heterogeneity for RNaseH sequence from each of the species. Estimated copy numbers and nuclear haploid DNA content are also shown, along with the reference for each estimate.

Species	Heterogeneity	Copy number	Reference	1C DNA/ pg	Reference
S. latifolia	0.278	4838.	Chapter 4	2.7	(Costich et al., 1991)
S. marizii	0.402	4633.	Chapter 4	3.6	Chapter 2
P. savitum	0.518	4000	(Pearce et al., 2000)	4.88	(Bennett and Smith, 1976)
V. faba	0.397	1000000	(Pearce et al., 1996a)	20.36	(Bennett et al., 1982)
P. abies	0.306	750000	(L'Homme et al., 2000)	18.6	(Siljak-Yakovlev et al., 2002)

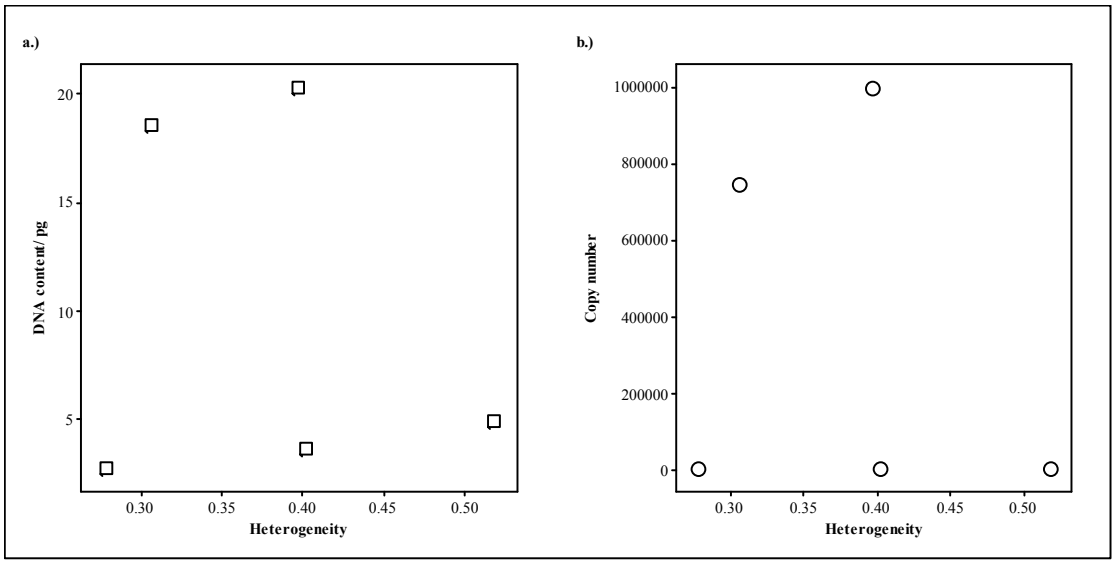


Figure 5.3: Relationship between within species heterogeneity and a: DNA content and b: copy number in the species studied.

5.4. Discussion

5.4.1. Sequence heterogeneity

Two results which stand out clearly from the phylogenetic analysis of the RNaseH sequence are, firstly, the lack of congruence between the sequence phylogeny and the evolutionary relationship between the species from which they derive and, secondly, the highly heterogeneous nature of the sequences examined, even from conspecific sequences. Together with the finding of strong purifying selection (discussed below), this is consistent with a purely vertical transmission of these elements, with a number of distinct and very ancient lineages of retrotransposons within all the species examined. Indeed, the phylogenetic analysis provides strong evidence that there are at least two major families of Ty1-*copia* retrotransposons present in *P. savitum* and *Silene* which originate prior to speciation between these two groups. This finding contrasts markedly with studies of retrotransposons in animals (e.g. Mount and Rubin, 1985), but is consistent with previous results from studies in plants, which appear to be much more tolerant of retrotransposon accumulation (Kumar and Bennetzen, 1999). This pattern has previously been reported from analyses of a number of sequences from plant Ty1-*copia* retrotransposons e.g. in the reverse transcriptase gene (Flavell et al., 1992b; Flavell et al., 1992a; Pearce et al., 1996a; Flavell et al., 1997; Matsuoka and Tsunewaki, 1999) and the U3 regulatory sequence (Vernhettes et al., 1998). All these studies have revealed high levels of sequence heterogeneity between conspecific retrotransposon sequences.

Estimates of heterogeneity rely on the assumption that retrotransposon sequences are randomly sampled from the wider population of retrotransposons. Work described in Chapter 4 suggests that this assumption may not be valid for these sequences, as PCR-selection may cause a bias towards elements with higher GC content at the primer binding site. This would be expected to result in a sample of sequences which are more closely related than would be seen given random sampling and are, therefore, less heterogeneous than random sequences from the underlying Ty1-*copia* population. This would be expected to affect any study using

degenerate primers and, as such, should be taken into account when interpreting the results of such studies.

5.4.2. Rates of substitution

The generally low values of K_a/K_s indicate that strong purifying selection is acting on most of the sequences examined. This finding is consistent with the regions status as a coding region, and suggests that the majority of the sequences included in the study represent active (or recently active) retrotransposon copies. Whilst a number of comparisons did not show a K_a/K_s value significantly different from 1, this was generally associated with highly similar sequences where the degree of sequence homogeneity would render reliable estimations of these parameters in the relatively short sequences impossible. This is confirmed by the generally high *p*-values associated with such comparisons.

5.4.3. Heterogeneity and copy number/ DNA content

Theoretical models suggested that the heterogeneity of retrotransposon sequence within a species should be correlated with the genomic copy number (Charlesworth, 1986). Such a relationship has previously been demonstrated between closely related species in *Vicia* (Pearce et al., 1996b). However, these results, from a number of distantly related species, do not show a similar pattern. In particular, the two species with vastly inflated copy numbers (*V. faba* and *P. abies*), do not show greater sequence heterogeneity than the other species.

When the relationship between DNA content and heterogeneity is examined there is, again, no overall relationship, but there is the suggestion of a trend towards positive correlations within large and small genomes. Given the uncertainty over the accuracy of copy number estimation in *Silene*, overall genome size may be a more accurate predictor of copy number, but with such few species, no firm conclusions can be drawn from these results.

Nevertheless, the theoretical model described in Charlesworth (1986) assumes equal transpositional and mutational rates, and hence a departure might not be unexpected

when considering such a wide diversity of retrotransposon populations from across a range of genera. Interestingly, sequence heterogeneity in *S. marizii* appears to be greater than that in *S. latifolia*, which is consistent with a greater copy number in the species if the theoretical assumptions are valid. Nevertheless, such a conclusion is highly speculative. More data from a variety of *Silene* species together with reliable estimates of Ty1-*copia* copy numbers would allow a better study of the relationship between copy number and heterogeneity within the genus.

5.4.4. Sequence divergence between *S. latifolia* and *S. marizii*

S. marizii and *S. latifolia* are very close relatives, hybridising readily, and with overlapping species distributions (Talavera, 1990). As such, it is surprising to find that there is such a clear phylogenetic distinction between RNaseH sequences from the two species. Indeed, the genetic distance between the two most similar sequences between the two species was no less than that found between *V. faba* and *P. savitum* and was substantially greater than those found within species. These comparisons are not strictly valid as the numbers of sequences sampled in each species are not equal but, nevertheless, sampling effort was greatest in the *Silene* species, and as such would be more likely to detect similar sequences. As such, the results do indicate that there is significant distinction between populations of Ty1-*copia* retrotransposons within *S. latifolia* and *S. marizii*. The sequences included in the study represent only a fraction of the Ty1-*copia* copies, and introgression between the species cannot be ruled out. However, these results do suggest that such an occurrence is not common and that there may be a significant degree of genetic isolation between the two species. This is consistent with patterns of germination and DNA content variation seen in hybrids (Chapter 3). In the similar *S. latifolia/ S. dioica* system, strong ecological separation is responsible for genetic isolation between the two species (despite extensive hybridisation) (Goulson and Jerrim, 1997). These results suggest that similar isolation exists between *S. marizii* and *S. latifolia*.

5.4.5. Summary

These results are consistent with previous phylogenetic analyses of Ty1-*copia* class retrotransposons in plants, showing that they are present in a highly diverse population, with distinct retrotransposon lineages present both within species, as well as shared between them.

These results also suggest that levels of gene flow between *S. latifolia* and *S. marizii* may be more limited than would be suspected, given their relatedness and overlapping species distributions. Nonetheless, this finding is consistent with the phenotypic distinction between the two species, and suggests that species boundaries may be maintained by strong selection against hybrids.

Chapter 6 The impact of measurement error on estimation of nuclear DNA content using flow-cytometry

6.1. Introduction

6.1.1. Estimating DNA content

There are only very few species for which complete genome sequence information is available and, as such, any attempt to quantify variation in DNA content within a wide variety of species or individuals must, by necessity, rely on indirect methods of estimation. Early attempts to quantify DNA content used bulk DNA extracted from a large number of cells. Estimates of DNA content were made by techniques such as chemical analysis (Schmidt and Thannhauser, 1945) or reassociation kinetics (Britten and Kohne, 1968). However, the interpretation of these techniques is difficult and prone to substantial error. A more recent approach is to estimate DNA content by examining single nuclei. There are two main variants to this approach. In the first approach, nuclei are stained using the Feulgen method, and the absorption of monochromatic light used to estimate DNA content. Error induced by irregularly shaped nuclei, or non-homogenous staining are minimised through the technique of microspectrophotometry (Deely, 1955). The second example of the single nuclei approach is that of flow-cytometry. In flow-cytometry, nuclei are stained with a DNA specific fluorochrome, generally propidium iodide (PI), and held in suspension. The suspension is passed in front of a laser beam and the fluorescence, and thereby the DNA content, of each nucleus is measured by an

array of photo detectors. An internal standard (a population of nuclei with known DNA content) is included, prepared and run with the sample in order to account for such factors as instrument drift, variation in dye intensity (Dolezel and Bartos, 2005) or inhibition of dye uptake caused by cell-bound plant compounds (discussed in more detail below). The results of such an analysis are usually presented as a frequency histogram of measured fluorescence. Thus, separate populations of nuclei are represented as a number of defined peaks on the histogram, including the standard nuclei, the diploid nuclei, and, possibly, tetraploid nuclei (from cells undergoing mitotic division) (Figure 6.1). The position of each of these peaks can be determined using an appropriate measure of central tendency (mean or mode) and the ratio of the sample to standard peak used to calculate an estimate of DNA content. The ease by which very large populations of nuclei can be analysed in a short period of time makes flow-cytometry an attractive method for estimating DNA content and hence there has been a shift towards this approach in recent years (Bennett and Leitch, 1995; Bennett et al., 2000).

6.1.2. Error in DNA content estimation

Since the first systematic attempts to quantify DNA content within a wide variety of organisms, a great many studies have reported significant variation in DNA both between and within species. Interspecific variation in DNA content has been well established, but the suggestion that large variations in DNA content exist within species (e.g. Laurie and Bennett, 1985; Vekemans et al., 1996; Rayburn et al., 2004; Smarda and Bures, 2006) or even within populations (Mowforth and Grime, 1989) has remained controversial (Greilhuber, 1998). In part, this is due to a lack of conclusive evidence for any mechanism by which such variation could be maintained within a species, given moderate levels of interbreeding (Greilhuber, 2005; Murray, 2005), but another significant criticism is that there may be the potential for significant experimental error or bias to influence estimates of DNA content (Greilhuber, 2005). Indeed, any study of DNA content variation relies on the accuracy of the method employed to measure DNA content, and a perennial problem associated with such measurements is that of distinguishing methodological or stoichiometric induced variation from that caused by real variation in DNA content. Error associated with estimation of DNA content will

always act to increase the estimates of DNA content variation within the group being studied, and as such, all estimates of variation within such groups must be considered overestimates to a greater or lesser extent.

In order to test the extent to which error in DNA content estimation might be responsible for intraspecific variation in DNA content, a number of studies have reassessed published examples of such variation, and have shown that in some cases the variation initially reported proves impossible to replicate (Greilhuber and Obermayer, 1997; Temsch and Greilhuber, 2000). Indeed, even when variation is repeatable, there is no guarantee that this is due to true genome size variability and not simply variation in an underlying inhibiting factor (Greilhuber, 1988b; Noirot et al., 2000; Price et al., 2000; Noirot et al., 2005).

Estimation of DNA content using both microspectrophotometry and flow-cytometry requires the comparison of stained nuclei of both the sample and a standard with a known DNA content. Fundamental to this method of estimation is the assumption that dye uptake in the DNA of both sample and standard is similar. Error induced by the inhibition of Feulgen staining caused by secondary plant metabolites in particular have been well documented (Greilhuber, 1988a, 1997), but even using propidium iodide based flow-cytometry there is some evidence that substantial error may still be possible.

6.1.3. Propidium iodide staining inhibitors

Propidium iodide based flow-cytometry is generally held to be a more robust method for estimating genome size than Feulgen microspectrophotometry (Greilhuber, 2005), but even using this method, a number of studies have found evidence that some cell bound compounds may inhibit the uptake of PI e.g. caffeoylquinic acid and caffeine in coffee trees (*Coffea sp*) (Noirot et al., 2000, 2002; Noirot et al., 2003), tannins in *Pinus* (Greilhuber, 1986, 1988b), or unidentified compounds in sunflower (*Helianthus annuus*) (Price et al., 2000). With internal standardisation, flow-cytometry should be robust to the effect of such error, unless the effect of staining inhibitors varies between sample and standard nuclei. Unfortunately, it appears that this is indeed the case in some instances, as

demonstrated by at least two studies (Price et al., 2000; Noirot et al., 2005) which showed that the concentration of endogenous staining inhibitors in the sample cytosol had a differential effect on the inhibition of PI uptake in standard and sample nuclei, with the magnitude of the difference depending on the concentration of cytosol present in the sample preparation. As such, estimates of DNA content would be expected to vary depending on the exact conditions of the sample preparation.

Fewer studies have examined the effect of random error in measuring the position of standard and sample peak positions. Indeed, the large number of nuclei that are sampled during a flow-cytometry run (numbering in the thousands) may suggest that such error will be negligible and, furthermore, the effect of such random events will not lead to systematic error in the manner that the stoichiometric effects described above will. Nevertheless, because DNA content estimates are based on the ratio of sample to standard peak position, when the fluorescence of the standard nuclei is small, even modest error in measuring the position of the standard peak may have large effects on the estimate of DNA content (Johnston et al., 1999).

For the reasons described above, it is always necessary to examine estimates of DNA content critically, and to question whether observed variation in DNA content (where detected) is real, or simply a consequence of experimental error. The aim of the current study is to test whether some of the basic assumptions of the flow-cytometric procedure that is used elsewhere in this thesis are valid, and what the effects of error or bias are likely to be for these data sets.

6.2. Methods

6.2.1. Estimating DNA content

The flow-cytometric method used to estimate DNA content in *Silene* (a detailed description of which is given in Costich et al (1991) and also in Chapter 2) compares the relative fluorescence of stained sample nuclei with that of nuclei from an internal standard of known DNA content. In the protocol described above, chicken red blood cells, with a known nuclear DNA content of 2.33pg, are used as the internal standard.

An idealised model of the procedure for estimating DNA content using internally standardised flow-cytometry can be written as:

$$\gamma_{ij} = g_{ij} = 2.33\left[\frac{\theta_i \varphi_{ij}}{\kappa \varphi_{ij}}\right],$$

where γ_{ij} is the true DNA content of the ith plant (in pg) (this can be assumed as constant across all values of j); g_{ij} is the estimated DNA content of the ith plant on the jth observation; θ_i is an idealised form of the mean fluorescence of sample nuclei (which is constant for all observations within the ith plant), κ is an idealised form of the mean fluorescence of the standard nuclei (which is constant in all observations) and φ_{ij} is a term that accounts for variation in the observed fluorescence caused by variation in sample preparation or instrument drift. Thus, $\theta_i\varphi_{ij}$ is the observed mean fluorescence of the sample nuclei (sample peak position, which can be simplified to 's'), and $\kappa\varphi_{ij}$ is the observed mean fluorescence of the standard nuclei (standard peak position, which can be simplified to 'c'). 2.33 is simply the nuclear DNA content (in pg) of the chicken standard. There are two clear features of this description: firstly, the estimate of DNA content is independent of the standard peak position and secondly, for any given value of DNA content, the ratio between sample and standard peak position is constant (i.e. there is a proportional relationship between the two). These two features may be considered the fundamental assumptions of the method.

6.2.2. Possible sources of error or bias

There are a number of mechanisms which might introduce error or bias into estimates of DNA content as outlined above. In the simplest case, error terms accounting for random variation in measuring the position of the sample and standard peak positions might be included. These error terms would be independent, with means of 0. As such, the estimate of DNA content would no longer be independent of the measured position of the sample peak (i.e. there would be a linear relationship between g and $1/c$), but there would be no bias in DNA content estimates, i.e.:

$$g_{ij} = 2.33\left[\frac{\theta_i\varphi_{ij} + x_{ij}}{\kappa\varphi_{ij} + y_{ij}}\right] = 2.33\left[\frac{s_{ij}}{c_{ij}}\right],$$

where x_{ij} and y_{ij} are errors associated with measuring the position of the sample and standard peak, respectively.

An alternative method by which error might influence the estimates of DNA content is that the measured fluorescence of either the standard or sample nuclei (or both) are affected by the action of cell-bound staining inhibitors. If the magnitude of the effect is consistent between standard and sample nuclei, then there will be no effect on the independence of g and c, and neither will there be any bias in the estimation of DNA content. However, if staining inhibitors have a differential effect on standard and sample nuclei, then again, the independence of g and c will be violated. Furthermore, because the absolute magnitude of the difference will depend on the concentration of the cytosolic compounds, there is also likely to be a significant deviation from the proportional relationship between the sample and standard peak position.

6.2.3. Test for the independence of the standard peak

In order to test for the presence and strength of a relationship between the estimates of DNA content (g) and the position of the standard peak (c), the outputs from 250 flow-cytometry runs were studied using a least-squares linear regression model to examine the relationship between g and $1/c$. The majority of these runs represented *S. marizii* samples, although some were *S. latifolia* and others represented

individuals from F1 or F2 generations in crosses between *S. marizii* and *S. latifolia*. Output from each of the flow-cytometry runs was assigned to one of four quality categories (poor, fair, good or excellent) in order to test whether subjective assessments of quality was a useful method of quality control when conducting analyses using flow-cytometry. Quality assignments were based on the shape and definition of the sample and standard peaks. Example histograms of these quality categories are shown in Figure 6.1.

6.2.4. Variation in DNA content estimates within plants

In addition to the analysis described above, a subset of 7 plants was selected for further analysis. These 7 plants consisted of 2 individuals of *S. latifolia*; 3 of *S. marizii* and 2 F1 hybrids. At least one male and one female were included from each of these three groups. Replicate measurements of DNA content were performed on each individual. An analysis of covariance (ANCOVA) was performed, using the software package Openstat (Miller, 2007), to examine the within and between plant variation in estimated DNA content, fitting the regression on $1/c$ as a covariate.

To test for the heterogeneity in the measured position of the standard peak between plants, a one-way ANOVA was performed on these data examining this effect. Two additional ANOVAs were performed examining the effect of plant on the measured position of the sample peak, and on the estimate of DNA content.

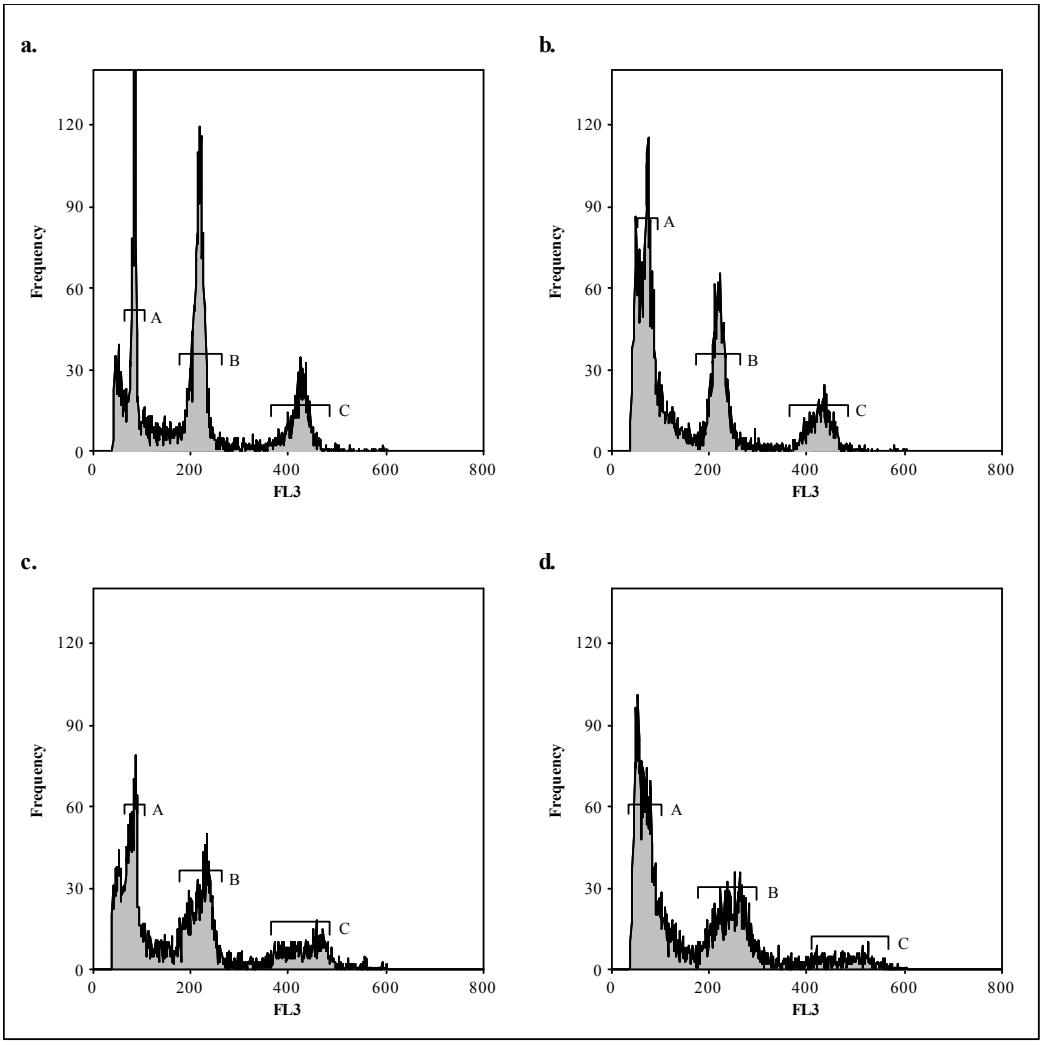


Figure 6.1: Examples of frequency histograms from each of the four quality categories (a: excellent, b: good, c: fair and d: poor). The y-axis shows the number of nuclei falling within each integer value of fluorescence (FL3, x-axis). Horizontal bars indicate the approximate location of populations of nuclei; A: standard nuclei, B: sample nuclei (diploid), and C: sample nuclei (tetraploid).

6.2.5. Test for proportionality between sample and standard peak position.

As described above, one of the fundamental assumptions of the flow-cytometric method is that there is a proportional relationship between the position of the sample peak and the position of the standard peak. If cell-bound staining inhibitors have a differential effect on each type of nucleus, then this assumption may be violated. A linear regression approach was adopted in order to test this assumption. Departure from a proportional relationship was tested by examining whether the calculated y-axis intercept of the linear regression was significantly different from zero.

Due to the presence of measurement error in determining the position of both the sample peak and standard peak positions, the method of ordinary least-squares regression will tend to underestimate slope parameters and overestimate intercepts in datasets such as these (Sokal and Rohlf, 1981), and is, therefore, inappropriate for testing the proportionality of the relationship between sample and standard peak positions. Instead, model II standardised major axis (SMA) linear regression models were fitted to each plant using the computer program SMATR (Falster et al., 2006), implementing the routines described in Warton (2006). Calculated intercepts were compared against the null hypothesis of 0 intercepts using a *t*-test.

One consequence of the presence of staining inhibitors, whatever their mode of operation, is that there will be a difference in the measured position of the standard peak when it is run alone, compared to when it is run in the presence of sample nuclei. In order to test for this, a paired *t*-test was conducted. Each flow-cytometry run was paired with a corresponding run in which only the standard nuclei were run. For each pair of runs, the nuclei were stained using the same dye preparation, and were run consecutively on the flow-cytometer.

6.3. Results

6.3.1. Relationship between *g* and *c*

There was strong evidence that the estimate of DNA content was dependent on the position of the standard peak. The regression accounted for over half of the total variance in DNA content when all quality categories were considered together (Table 6.1). Within quality categories, regression models were also highly significant. The proportion of the total variance in the estimated DNA content explained by the regression model varied between quality categories, with regression models in the lower quality categories accounting for a greater proportion of the total variance than those in the higher quality assignments (Table 6.1, Figure 6.2).

Table 6.1: Regression parameters and significance values for the four quality categories and for the data set as a whole.

Category	N	Regression Equation	R^2	*F*-ratio	*p*
Excellent	80	251+3.68/c	33.5	39.2	<0.001
Good	85	362+2.31/c	43.8	64.8	<0.001
Fair	60	381+1.95/c	45.7	48.76	<0.001
Poor	25	407+1.63/c	76	72.9	<0.001
Overall	250	359+2.31/c	50.4	252	<0.001

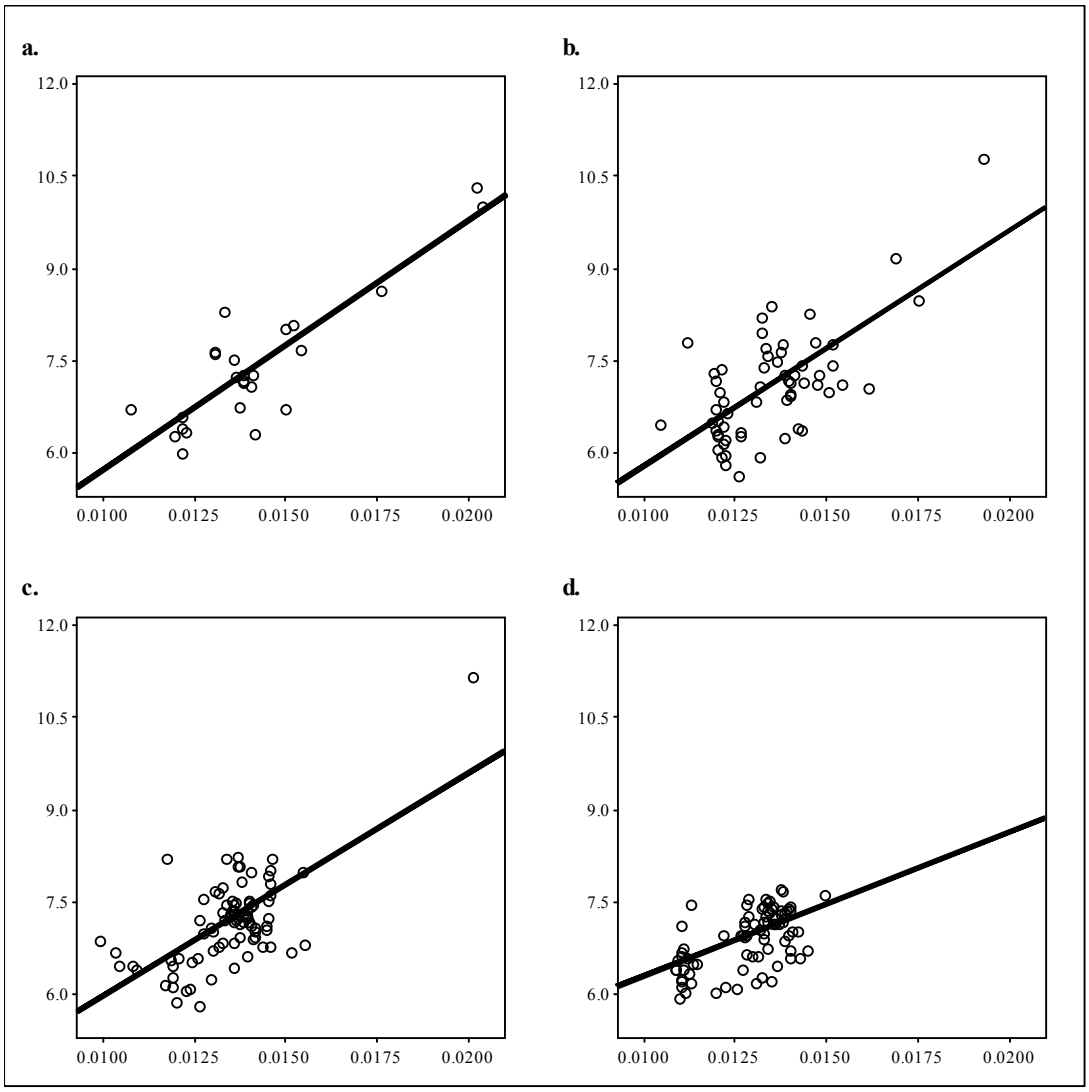


Figure 6.2: DNA content estimates in pg (y-axis) plotted against the inverse of the standard peak position (1/c) (x-axis). Each graph represents results from each of the four quality categories, (a: poor; b: fair; c: good; d: excellent). Solid lines in each graph show fitted linear regressions.

6.3.2. Within plant variation in DNA content estimates

There was no evidence for heterogeneity of within plant regression coefficients ($F_{6,77}=0.881$, $p=0.513$) and as such, the method of ANCOVA can be considered an appropriate method of analysis for these data.

The results of the ANCOVA analysis demonstrate that the regression on the inverse of the standard peak position and the effect of plant were both highly significant. The within plant variation in estimated DNA content (error mean square) (accounting for the regression on $1/c$) was substantially higher than the between plant variance in estimated DNA content (Plant mean square) (Table 6.2).

Table 6.2: ANCOVA table showing the effect of plant on estimates of DNA content, whilst accounting for the relationship between estimates of DNA content and standard peak position.

Source	Degrees of freedom	Sum of squares	Adjusted sum of squares	Adjusted mean square	*F*-ratio	*p*
Regression	1	4.5648	4.9011	4.9011	37.7	<0.001
Plant	6	6.4104	6.4104	1.0684	8.22	<0.001
Error	83	10.7914	10.7914	0.13		
Total	90	21.7666				

There was no evidence for a plant effect on the position of the standard peak, but there was very strong evidence of a plant effect on both the position of the sample peak and the estimate of DNA content (Table 6.3).

Table 6.3: Within plant means for the standard peak position, the sample peak position and the estimated genome size. The lowest row gives the results of a one-way ANOVA examining the effect of plant on the corresponding variable.

Sample	N	Sex	$\bar{c}$	$\bar{s}$	$\bar{g}$
C 7-1	15	F	74.33	218.9	6.87
C 7-2	14	M	73.19	203.41	6.48
C 78-01	12	F	74.39	220.78	6.94
C 8-1	12	M	73.22	231.92	7.39
C 8-2	10	F	73.64	225.33	7.14
C8-3	14	F	74.164	225.48	7.09
87-03	14	M	75.04	224.49	7.00
F (p)	91	-	0.35 (0.909)	7.48 (<0.001)	5.42 (<0.001)

6.3.3. Test for proportionality

Two of the plants examined showed evidence for an intercept that was significantly different from 0. These were individuals C78-01 and C87-03 (Table 6.4, Figure 6.3)

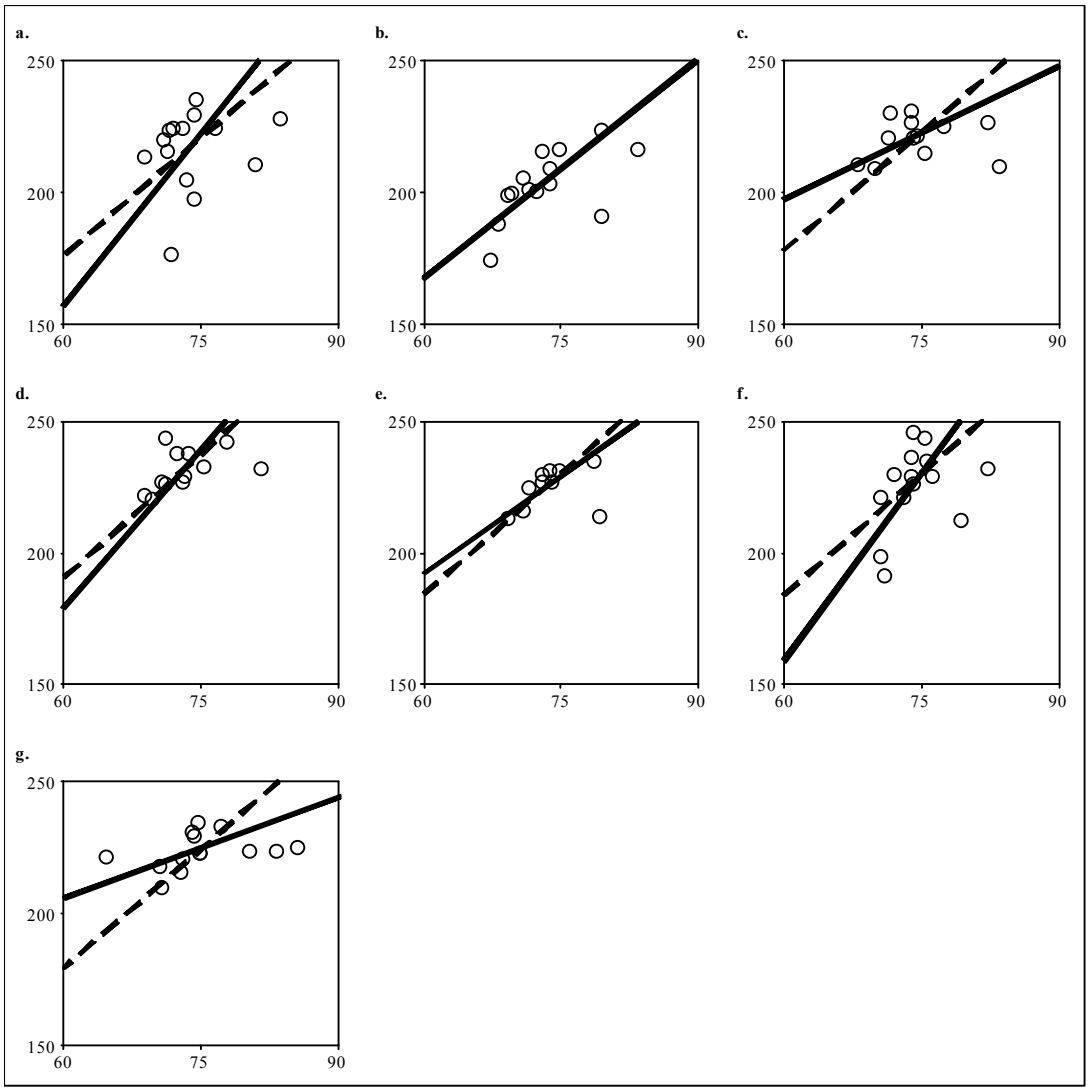


Figure 6.3: Standard (X axis) and sample (Y axis) peak positions for flow-cytometry runs within each plant (a: C7-1; b: C7-2; c: C78-01; d: C8-1; e: C8-2; f: C8-3 and g: C87-03). Black lines show the fitted model II linear regression. Dashed lines show fitted linear regressions, constrained to pass through the origin for comparison.

Table 6.4: Model II (SMA) regression parameters examining the relationship between standard peak position and sample peak position. The statistical significance of the intercept is indicated by its associated *t*-statistic and *p*-value.

Plant	N	Slope	Intercept	*t*	*p*
C7-1	15	4.38	-106.29	-1.23	0.240
C7-2	14	2.80	-1.37	-0.03	0.977
C78-01	12	1.69	95.25	2.40	0.037
C8-1	12	-55.69	3.93	1.75	0.111
C8-2	10	43.51	2.47	0.71	0.495
C8-3	14	4.70	-122.85	-1.32	0.212
C87-03	14	1.30	127.29	4.81	<0.001
Overall	91	3.50	-37.69	-1.44	0.155

6.3.4. Paired *t*-test

There was strong evidence that the measured position of the standard peak in sample runs was greater than the measured position of the standard peak in paired runs, in which only standard nuclei were run, although the differences were generally small (Table 6.5, Figure 6.4). The directions and magnitude of these differences seem generally consistent between plants. It appears that the effect is caused only by changes to the shape of the standard peak as there is no evidence for a disparity when mode is used as the measure of standard peak position (Table 6.5, Figure 6.4).

Table 6.5: Means and standard deviations for paired sample/standard minus standard only runs using either the mean or mode of the standard peak as a measure of peak position. *t*-values and statistical significance are also shown.

	Mean of standard peak				Mode of standard peak			
Plant	Difference	Std Dev	*t*	*p*	Difference	Std Dev	*t*	*p*
C7-1	1.008	1.414	2.57	0.025	-0.31	4.57	-0.24	0.812
C7-2	-0.7	1.833	-1.32	0.213	0	2.256	0.00	1.000
78-01	1.111	2.421	1.38	0.206	-0.33	6.28	-0.16	0.878
C 8-1	0.791	1.936	1.36	0.205	0.909	1.375	2.19	0.053
C 8-2	0.39	1.513	0.82	0.436	0.7	1.636	1.35	0.209
C8-3	0.438	1.662	0.95	0.36	-0.462	3.55	-0.47	0.648
87-03	1.182	1.564	2.51	0.031	2.27	4.54	1.66	0.128
Overall	0.621	1.824	3.05	0.003	0.425	3.748	1.01	0.314

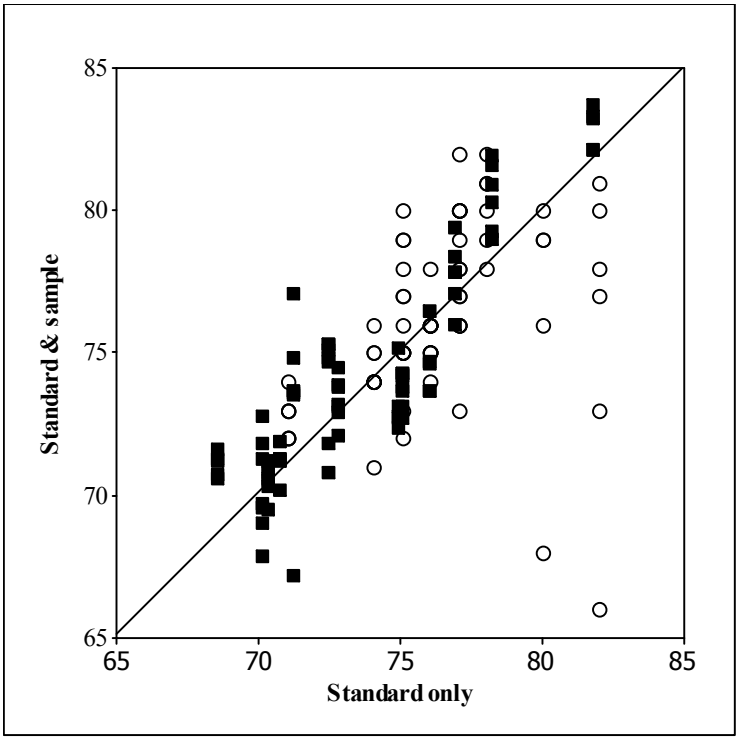


Figure 6.4: Paired measures of the position of the standard peak in runs using standard only and runs using sample plus standard. Open circles represent standard peak modes, and filled squares, sample means.

6.4. Discussion

6.4.1. Nature of error

Both the overall analysis of flow-cytometry runs and the within plant analysis of DNA content showed strong evidence for a relationship between the estimates of DNA content and the measured position of the standard peak. The highly significant relationship between the position of the standard peak and the estimate of DNA content does not, in itself, give any reason to be concerned with the method used. Indeed, given any degree of experimental error in measuring the standard peak position, such a relationship would be expected. Nevertheless, it is worrying, given the extent of variation in the estimates of DNA content, that such a large proportion of this variation can be attributed simply to variation in the position of the sample peak. This is open to two interpretations:

1. Significant error may exist in the measurement of the position of the standard peak.
2. Cytosolic compounds may be inhibiting PI uptake by DNA, and the effect is not consistent between sample and standard nuclei.

If the second interpretation were correct, then one would expect a significant departure from the proportional relationship between sample and standard peak positions within individuals. However, when replicate measurements on single individuals are considered, there is only evidence for a significant departure from proportionality in two cases. This suggests that most of the relationship between standard peak position and DNA content estimates can be explained simply by random error in the measurement of the standard peak, with no need to invoke the existence of staining inhibitors to explain the results. Nevertheless, the two examples where there is evidence against proportionality cannot be discounted and the possibility remains that such factors may be affecting estimates of DNA content in at least some plants.

That a significant proportion of the variability in DNA content estimates are simply a result of error in measuring the standard peak position seems at odds with the apparently high definition of the standard peak, and the low variance in the fluorescence of the standard nuclei in almost every flow-cytometer run examined. However, the nuclear DNA content of chicken red blood is low compared to the nuclear DNA content of the plants being examined, and as such small error associated with the standard peak position may have large effects on the estimate of nuclear DNA content, an effect previously suggested by Johnston et al (1999). Indeed, if the mean standard peak position is assumed to be approximately 75 fluorescence units, and the mean sample peak position is 208, then with similar absolute error in measuring each of the peak positions, the error in measuring the standard position will have almost three times the effect on estimated DNA content than error in measuring the sample peak.

6.4.2. Adjusting estimates of DNA content variation

These results suggest that error in measuring the position of the standard peak may be a significant source of error in this method of DNA content estimation, which will act to increase estimates of DNA content in survey populations. As such, the method of linear regression may allow such estimates to be corrected by fitting a linear model accounting for the relationship between g and $1/c$. The residual variance can be taken as a new estimate of the variance in DNA content. This is

still likely to be elevated relative to the true value, but will provide a more accurate indication of the degree to which DNA content varies in the individuals being studied.

These results do not suggest that the method for estimating DNA content is biased and does not cast doubt on the correlations observed between DNA content and phenotypic traits in the genus. However, they do suggest that absolute values for levels of DNA content variation should be treated with caution and that it is important to apply some form of quality assessment when conducting flow-cytometry procedures. The results suggest that fitting a linear regression model to the data may allow some degree of correction to be applied to the estimated levels of DNA content variation within the experimental population. This method has the advantage of being applicable to any dataset without the need for conducting replicate measurements on individuals. However it must be borne in mind that this can only correct for error in estimating the position of the standard peak; error associated with measuring the sample peak will be impossible to estimate without replicate measurements. Nevertheless, if error associated with measurement of the sample peak position is of a similar magnitude to that associated with measuring the standard peak position then its effect on the estimate of DNA content will be much lower. Indeed, the residual variation in DNA content appears generally much more consistent between quality categories than the variation explained by the regression model.

6.4.3. Presence of inhibitors

The results of the comparison of standard only, compared to standard plus sample runs, indicated that there is a significant (if slight) increase in the measured position of the standard peak when it is run alone. This would appear to support the idea that cell-bound compounds inhibit dye uptake in these nuclei. However, if the *t*-test is performed again using the mode of the standard peak rather than the mean, this effect disappears. This result suggests that the effect of the sample on standard nuclei can be explained as an effect on the shape of the peak rather than on its position. This interpretation would be more consistent with fluorescent particles present in the sample preparation occupying a similar position to the standard

nuclei. Indeed, taken together these results suggest that as well as there being no evidence for a differential inhibition (between sample and standard), there is no evidence for the presence of any type of inhibitor at all (on the standard nuclei at least).

In addition, the results of the one-way ANOVA examining the effect of plant on the measured position of the sample peak suggest that there is no variation between plants in their effect on the standard peak position. This is reassuring as it suggests that even if there is some systematic effect of the sample on the measured position of the standard peak then it is very unlikely to cause repeatable variation between either individuals or species in estimates of DNA content. As such, observed variation in DNA content between sexes, populations or species, are unlikely to be caused by bias in the method of DNA content estimation.

6.4.4. Within plant variation in DNA content

Again, the results suggest that a significant proportion of the variation in DNA content may be explained by a regression on the inverse of the standard peak position. This is consistent with the results from the larger analysis described above. The strength of this relationship appears to be consistent within plants. When the relationship between the position of the standard peak and the estimate of DNA content has been accounted for, the within plant variance in the estimate of DNA content is modest. Again, this suggests that error in measuring the position of the standard peak may be the primary source of error in these data.

6.4.5. Internal standardisation

The method of internal standardisation employed here is open to the criticism that it is not true internal standardisation (Johnston et al., 1999; Dolezel and Bartos, 2005) because the standard nuclei are stained prior to their incubation with the sample preparation. Although this limits the ability for this study to detect the effects of staining inhibitors that affect only the sample nuclei, the strong effect of the standard peak on estimates of DNA content cannot be discounted. Nevertheless, a

study using a co-prepared standard would be useful to fully test for the presence of cytosolic staining inhibitors.

6.4.6. Summary

The results from this analysis suggest that whilst there may be significant error in the estimate of DNA content, there is no evidence to suggest that differences in DNA content between individuals are purely due to experimental error, or that estimates of DNA content are affected by cytosolic staining inhibitors. In particular, the consistent variation observed between species or sexes appears to be robust to the effects of experimental error. Nevertheless, the fact that such a significant proportion of the (relatively large) total variance in DNA content, observed in this study, may simply be due to error in the measurement of the standard peak position suggests that careful attention to the flow-cytometric protocol employed may be warranted and provide significant improvements in the accuracy of DNA content estimation within the genus.

These results emphasise the need for estimates of DNA content to be examined with a critical eye. Furthermore, they illustrate that subjective assessments by the operator can be an important method of quality control, and that, where possible, results that appear to be of low quality should be discarded, and the individual reanalysed.

Within this thesis, and in other studies within the *Silene* genus, the flow-cytometric technique examined here is used to detect relationships between DNA content and other phenotypic traits, rather than to detect absolute values of DNA content variation. As such, the fact that estimates of DNA content are unlikely to be biased is more significant that the fact that estimates of DNA content variation may be inflated. However, large error in the estimation of DNA content may reduce the potential to detect such relationships, particularly if they are weak. As such, reducing the extent of error inherent in the method would be of considerable interest. One suggestion is that the low DNA content of the standard nuclei causes that measurement error associated with the standard to have a large effect on the estimate of DNA content. An alternative internal standard more closely matched to

the sample nuclei may significantly improve the accuracy of DNA contents. This, in turn may improve the ability to detect relationships between DNA content and floral traits.

Although absolute values of DNA content variation are of secondary importance, the degree of variation in DNA content variation within the genus is, nonetheless, of considerable interest. These results suggest that a linear regression on the inverse of the standard peak position may allow the estimates of DNA content variation to be adjusted to account for, at least, some of the variation induced by experimental error.

Chapter 7 General conclusions and future work

The overall objective of the work described in this thesis was to investigate patterns of DNA content variation and its phenotypic consequences in *Silene*. This was achieved by using a comparative approach between two closely related species in the genus: *S. marizii* and *S. latifolia*.

7.1. The relationship between DNA content and flower size

The survey of phenotypic variation in *S. marizii* has revealed that the genetic basis for variation in flower size is likely to be different in this species from that found in *S. latifolia*. In particular, the phenotypic correlations between DNA content and floral morphology appear to be either much weaker or absent, and there appears to be a much greater potential for independent, sex-specific evolution between the three measures of flower size used in these studies. The study of F2 hybrids appears to support this and, furthermore, the lack of correlation between DNA content and flower size in these hybrids suggests that the cause of the interspecific association between DNA content and flower size may be different in nature to the mechanism that causes the same relationship within *S. latifolia*. In particular, it seems likely that either a significant part of the interspecific differences in DNA content may be independent of flower size, or that genetic mechanisms other than DNA content are a significant cause of variation in flower size between the two species. As such, hybrids between these species may not provide a useful tool for studying the mechanisms by which DNA content might influence phenotype in these species, although the nature of interspecific differences in these traits are of considerable

interest in their own right, particularly the evolutionary mechanisms by which large scale variation in DNA content can arise between two closely related species.

7.2. Interspecific differences in DNA content

One striking result to come from the survey of phenotypic variation in *S. marizii* is the lack of consistent sexual dimorphism in DNA content, particularly given that it is such a predominant feature of *S. latifolia*. Such a result must, necessarily, be interpreted in terms of sex-chromosome heteromorphy. Why then does *S. marizii* seem to have avoided the evolution of heteromorphic sex-chromosomes when its close relative, *S. latifolia*, has not? The survey of hybrid populations suggests that the difference can be interpreted as an enlarged X-chromosome in *S. marizii* (relative to that of *S. latifolia*) rather than a smaller Y-chromosome, although this requires verification. Whether this is caused by greater recombination between *S. marizii* sex-chromosomes, or independent evolution of X-chromosomes is a question that remains to be addressed. In particular, cytological studies of sex-chromosomes would be particularly valuable in establishing the extent of heteromorphy in sex-chromosomes within each species.

7.3. Y-chromosome polymorphism in *S. marizii*

One of the most intriguing results to come out of this work is the suggestion that sexual dimorphism in DNA content may act in opposite directions between *S. marizii* populations. The cause of such a sexual dimorphism necessarily rests with the Y-chromosome in this species. This finding raises a number of interesting questions regarding the mechanisms by which such a polymorphism might be maintained within a population. Surveys of DNA content within a wider range of *S. marizii* populations would be useful to establish the true extent of such polymorphisms within the species, whilst comparisons between *S. marizii* and *S. latifolia* sex-chromosomes using either cytogenetic techniques, or molecular markers would test whether these polymorphisms evolved within the species or whether they are a result of introgression from *S. latifolia*.

7.4. DNA content in hybrids

The work with *S. marizii*/ *S. latifolia* hybrids suggests that there may be unusual patterns of DNA content segregation during such crosses. In particular, there is the suggestion that variation in DNA content is similar in each generation and that there may be an upward trend in DNA content between successive generations of an F2 crossing design. This would suggest that hybrid populations are biased with regard to their parents. This is consistent with significant levels of reproductive isolation between the species that is likely to exist given the overlapping species distributions, and the ease by which hybridisation occurs, and suggests that variation in DNA content may play a role in maintaining species boundaries. Further investigations of patterns of DNA content segregation using a greater number of parental crosses would be useful in order to test the generality of this result.

7.5. Rates of germination and sex-ratio bias

Low germination rates and low rates of flowering seem to be common amongst F2 hybrids between *S. latifolia* and *S. marizii*. To a large extent, this can be attributed to variation in seed weight, although there is no evidence to suggest seed weight is related to either DNA content or flower size.Nevertheless, low rates of germination and flowering may present a practical barrier to the use of such hybrids to study the genomic basis, and correlated phenotypic effects, of DNA content variation within *Silene*. If low germination rates are environmentally induced, then it may be possible to improve rates by altering germination conditions, although a number of germination strategies trialled during this work proved ineffective at improving rates of germination.

Another common feature of the work both with *S. marizii* and with hybrid populations was biases in sex-ratios, with an excess of males in populations of *S. marizii* and an excess of females in hybrid populations, though only when *S. marizii* was the paternal plant. Low rates of germination mean that it is impossible to tell at which point these biases are introduced or whether they would be a feature of natural populations. Nevertheless, the pattern of sex-ratios seen in hybrid

populations is consistent with the mechanism of sex-ratio distortion previously described in *S. latifolia* and also with the theory that such a mechanism of sex-ratio distortion is absent in *S. marizii*. If further studies of hybrid populations confirmed this result, then the *S. marizii*/ *S. latifolia* system would provide a valuable tool for studying the extent and evolution of sex-ratio distorters. In particular, it would allow the true extent of such distorter alleles to be assessed in a genetic background which lacked a male restorer allele.

7.6. Ty1-*copia* retrotransposons in *Silene*

Retrotransposons are significant constituents of plant genomes, and furthermore, may have direct impacts on phenotype. The studies of Ty1-*copia* class retrotransposons described in this thesis demonstrate that they are present in high numbers and as a highly heterogeneous population. The phylogenetic study also suggests a relatively high degree of genetic isolation between *S. latifolia* and *S. marizii*, a result that is consistent with the morphometric and DNA content distinction between the species.

The degree of variation in copy number indicated by the SSAP based method, particularly within populations and species, seems improbable given moderate levels of interbreeding and the low levels of transpositional activity that are characteristic of these elements. As such, it seems likely that there is substantial error associated with the method, and that it may not be an effective approach for assaying the contribution of these elements to DNA content in large survey populations. Therefore, an alternative method may be required to investigate the contribution of specific genetic elements to DNA content variation and the associated variation in flower size within *Silene*.

7.7. Flow-cytometry methods

The results from the quality assessment of the flow-cytometric method used in this (and other) studies of DNA content variation within *Silene*, underline the importance of a critical assessment of such DNA content estimates. The results of this study do not give any reason to conclude that estimates of DNA content are

biased, or that the action of cytosolic staining inhibitors within the samples may be affecting such estimates, but they do suggest that error associated with estimating DNA content may represent a significant proportion of the variance in such estimates . In particular, they suggest that much of the variation in DNA content estimates may be caused by error associated with the measurement of the standard peak position.

This may, in part, be due to the low nuclear DNA content of the standard nuclei relative to that of the sample nuclei. This suggests that substantial improvements in the accuracy of DNA content estimates may be possible by using an alternative flow-cytometric protocol, in particular by using an internal standard that is better matched to the sample. Increasing the accuracy of DNA content estimates would allow a much better assessment to be made of the extent of variation in DNA content, both within and between species. It would also allow weaker correlations between DNA and phenotype to be established, the nature and extent of sexual dimorphism in DNA content to be determined with more accuracy, and the pattern of DNA content segregation during hybridisation to be assessed more precisely. As such, improving the accuracy of DNA content estimation must be considered a high priority by future studies.

7.8. Conclusion

The work described in this thesis demonstrates that an association between DNA content and flower size is not a ubiquitous feature within species of the section *Elisanthe.* Indeed, these results demonstrate that the direction of sexual dimorphism in DNA content may be independent of the direction of sexual dimorphism in flower size. Furthermore, they suggest that interspecific variation in DNA content and flower size may have a very different basis to variation in these characters within species. . Nevertheless, differences in DNA content between *S. latifolia* and *S. marizii* may provide important information about differentiation between the two species, particularly with regard to the evolution of heteromorphic sex-chromosomes. The results also suggest that there is a substantial degree of genetic isolation between the two species that may be independent of ecological isolation. Biases observed in the DNA content of hybrid populations suggest that large scale

differences in DNA content may play a role in the maintenance of species boundaries within the genus. Fundamental to much of the work described here (and previous studies of DNA content variation within the genus) is the flow-cytometric method. These results suggest that this method may be prone to substantial errors. Whilst this is concerning, the potential to improve the accuracy of this method exists. If this can be achieved, future studies will benefit greatly from improved accuracy in estimates of DNA content.

Overall, the work described in this thesis demonstrates that, whilst DNA content variation within *Silene* is considerable and correlations between flower size and DNA content exist both within and between species, the genetic control of floral variation (and sexual dimorphism in particular) may be due to a number of factors other than DNA content. The results also suggest that differences in DNA content, observed between species, may be a result of factors other than selection on flower size.

Appendix A Partial RNaseH sequences used in the thesis

In this appendix, the source and nature of partial RNaseH DNA sequences used elsewhere in this thesis are given. Where sequences were retrieved from previous publications, the EMBL nucleotide database accession number is provided, whilst sequences produced during the work described in this thesis are identified by a unique number.

Table A.1: List of partial RNaseH sequences retrieved from the EMBL nucleotide sequence database that use motif 1 as the 5′ primer binding site. Sequence lengths are indicated along with the number of G or C bases at RNaseH motif 2. All sequences are from Pearce (2007).

EMBL Accession/ ID	Species	Sequence length/ bp	Motif 2 GC bases
AM408587	*Vicia faba*	208	6
AM408586	*Vicia faba*	209	6
AM408585	*Vicia faba*	209	6
AM408584	*Vicia faba*	210	5
AM408583	*Pisum sativum*	208	6
AM408582	*Pisum sativum*	208	7
AM408581	*Pisum sativum*	208	5
AM408580	*Pisum sativum*	205	5
AM408579	*Glycine max*	208	4
AM408578	*Glycine max*	208	5
AM408577	*Glycine max*	209	5
AM408576	*Glycine max*	208	4
AM408575	*Glycine max*	208	5
AM408574	*Glycine max*	208	5

Table A.2: List of partial RNaseH sequences retrieved from the EMBL nucleotide sequence database that use motif 2 as the 5′ primer binding site. Sequence lengths are indicated along with the number of G or C bases at RNaseH motif 2. All sequences are from Pearce et al (1999).

EMBL Accession/ ID	Species	Sequence length/ bp	Motif 2 GC bases
AJ243035	*Pisum Savitum*	165	6
AJ243036	*Pisum savitum*	238	9
AJ243037	*Pisum savitum*	229	7
AJ243038	*Pisum savitum*	162	7
AJ243039	*Pisum savitum*	206	7
AJ243040	*Pisum savitum*	249	8
AJ243041	*Pisum savitum*	249	9
AJ243042	*Pisum savitum*	274	8
AJ243043	*Pisum savitum*	200	7
AJ243099	*Vicia faba*	138	7
AJ243100	*Vicia faba*	175	8
AJ243101	*Vicia faba*	134	5
AJ243102	*Vicia faba*	241	7
AJ243103	*Vicia faba*	239	5
AJ243104	*Vicia faba*	235	6
AJ243312	*Picea abies*	186	8
AJ243313	*Picea abies*	163	8
AJ243314	*Picea abies*	292	5
AJ243315	*Picea abies*	171	6
AJ243316	*Picea abies*	249	6
AJ243317	*Picea abies*	170	8
AJ243318	*Picea abies*	177	7
AJ243319	*Picea abies*	146	5

Table A.3: List of partial RNaseH sequences generated during work described in this thesis that use motif 2 as the 5′ primer binding site. Sequence lengths are indicated along with the number of G or C bases at RNaseH motif 2.

EMBL Accession/ ID	Species	Sequence length/ bp	Motif 2 GC bases
1.1	*Silene latifolia*	357	7
2.19	*Silene latifolia*	285	7
3.16	*Silene latifolia*	375	7
4.12	*Silene latifolia*	389	8
4.8	*Silene latifolia*	437	9
5.35	*Silene marizii*	340	9
5.42	*Silene marizii*	338	9
6.12.4	*Silene latifolia*	209	8
6.16.4	*Silene marizii*	353	10
6.19.4	*Silene latifolia*	229	8
6.24.4	*Silene latifolia*	220	8
7.13.5	*Silene latifolia*	134	9
7.7.6	*Silene latifolia*	230	9
7.9.6	*Silene latifolia*	295	9
7.12.6	*Silene latifolia*	182	9
7.22.6	*Silene latifolia*	255	8
9.1.4	*Silene latifolia*	226	7
9.1.21	*Silene latifolia*	82	8
9.2.4	*Silene marizii*	252	7
9.2.21	*Silene marizii*	253	7
11.1	*Silene latifolia*	176	10
11.2	*Silene latifolia*	139	8
11.11	*Silene latifolia*	140	8
11.17	*Silene marizii*	172	8
11.18	*Silene marizii*	234	10
11.22	*Silene marizii*	236	8

Appendix B LTR retrotransposon based markers

B.1. Introduction

One significant barrier to genomic analysis in *Silene* species is the lack of co-dominant molecular markers with which to construct a genetic map. One potential resource for generating co-dominant molecular markers in plant species is retrotransposons. There are several types of retrotransposon based molecular marker systems. The first is an AFLP-like technique known as site-specific amplification polymorphism (SSAP) (Waugh et al., 1997). Like AFLP, this is a dominant marker system, and is a useful method for rapidly generating large numbers of polymorphic loci (Flavell et al., 1998), but contrary to AFLP, which uses two random primer binding sites, SSAP uses one random, and one site-specific primer. The second type of retrotransposon based marker system is retrotransposon based insertion polymorphism (RBIP) (Flavell et al., 1998). The RBIP marker system identifies the presence or absence of specific retrotransposon insertions. Because retrotransposons can be many thousands of base pairs in length, both allelic states cannot be detected by a single pair of PCR primers (as in comparable systems such as microsatellites). As such, a third primer must be added, which is internal to the retrotransposon. This is illustrated graphically in Figure B.1. Both marker systems require sequence information from the LTR region of the retrotransposon. In addition, RBIP markers require flanking sequence from both ends of individual retrotransposon insertion sites. Retrotransposon based marker systems have a number of advantages over similar

systems, with high copy numbers in most plant genomes, a genome wide dispersal, and a high stability once present in an insertion site (Kumar and Hirochika, 2001).

In order to develop RBIP markers, LTR sequence data along with flanking sequence data, from either side of the insertion site, is required. LTRs are identical at each end of an LTR retrotransposon. This means that after 3′ LTR sequence has been isolated the same sequence can then be used to isolate 5′ flanking sequence by amplifying outwards from the 5′ LTR. Isolation of 3′ flanking sequence is achieved by amplifying from the 5′ flanking sequence (back towards the insertion site) in an accession lacking the insert.

The work described in this Appendix describes attempts to develop RBIP molecular markers in *Silene*. Using the method described in Pearce et al (1999) to isolate LTR sequence from Ty1-*copia* class retrotransposons, a general overview of which is given in Figure B.2.

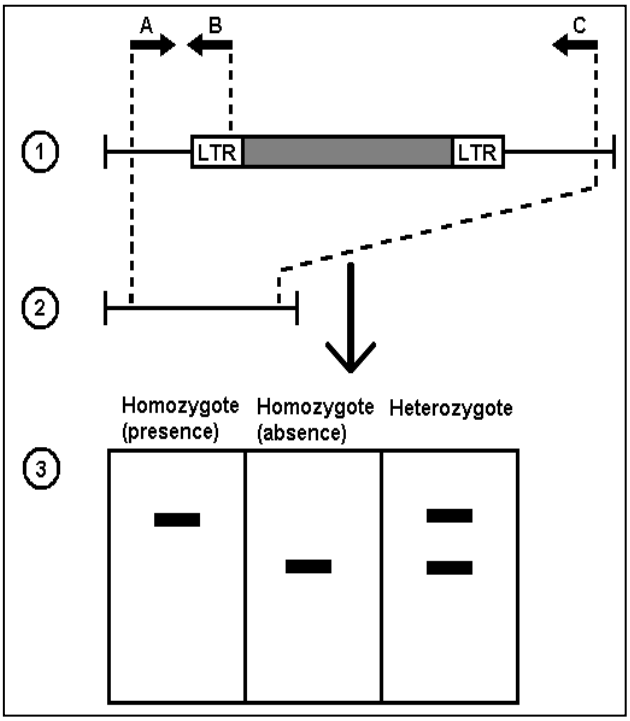


Figure B.1: Schematic showing an insertion site with (1) and without (2) an LTR retrotransposon. The three primer sites (A, B and C) are shown, along with the resultant banding pattern for the three genotypes (3).

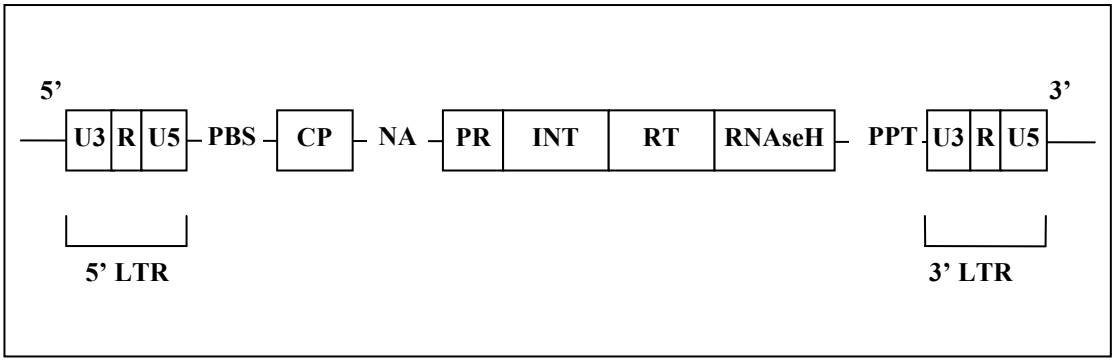


Figure B.2: General structural features of the Ty1-*copia* class of LTR retrotransposons. Within the LTR are the U3, R and U5, regions that contain signals for initiation and termination of transcription. The genes inside the retrotransposons code for capsid like proteins (CP), protease (PR), integrase (INT), reverse transcriptase (RT) and RNaseH. Also shown is a polypurine tract (PPT). The drawing is not to scale; LTR retrotransposons are highly variable in length, ranging from a few kbp to several thousand kbp.

B.2. Methods

Genetic material for the development of molecular markers was taken from the same individuals that were used to generate interspecific hybrids (Chapter 3). The isolation of LTR sequence data was achieved using a method adapted from that described in Pearce et al (1999). An overview of these steps is shown schematically in Figure B.3.

B.2.1.Digestion of genomic DNA

2.5µg of genomic DNA was digested with 1 unit of MseI, (New England Biolabs) in 100µl of restriction buffer 3 (New England Biolabs) at 37°C. 20µl aliquots of the reaction were removed after 2, 5, 10, 20 and 40 minutes. The restriction enzyme in these aliquots was inactivated by incubation at 80°C for 10 minutes. 5 µl of each aliquot was visualised by agarose gel electrophoresis. The aliquot that showed a mean fragment size closest to 1kbp was selected for adapter ligation. The sample was purified using a qiaquick PCR purification column according to the manufacturer's instructions.

B.2.2.Preparation of template DNA

Adapters were prepared by combining two single stranded oligos (MseI: 5'-GACGATGGATCCTGAG-3'; MseI2: 5'-TACTCAGGATCCAT-3'). Template DNA was prepared by ligating 2pmol of adapter to the digested DNA in ligation buffer containing 1mM ATP and 0.1 units of T4 DNA ligase. Reactions were incubated at 4°C for 16hrs. Ligation reactions were purified in a qiaquick PCR purification (Qiagen) column according to the manufacturer's instructions.

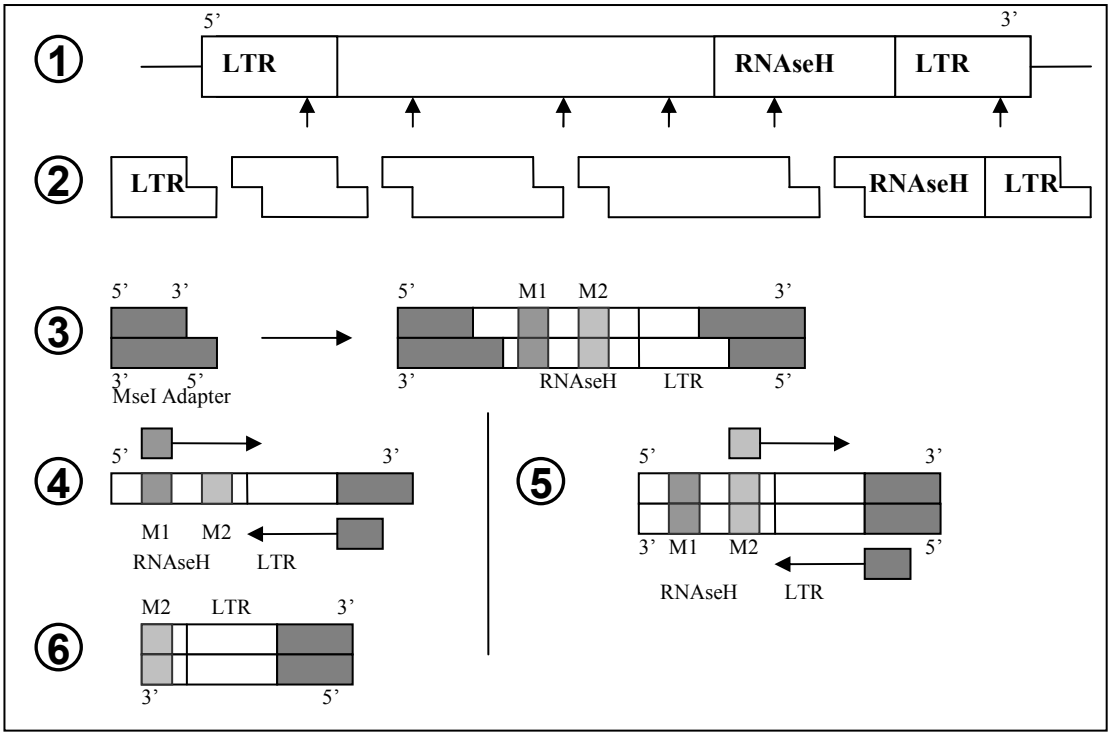


Figure B.3: Schematic diagram showing the steps taken to isolate LTR sequence from Ty1-*copia* retrotransposons. 1. Genomic DNA, showing a LTR retrotransposon insertion; arrows indicate cut sites for the MseI restriction enzyme. 2. Partial restriction digest with MseI. 3. Unphosphorylated MseI adapters are ligated to restriction fragments. 4. 1st PCR, using RNaseH motif 1 (M1) primer and a MseI adapter specific primer. 5. 2nd PCR with RNaseH motif 2 (M2) primer and the MseI adapter primer. 6. PCR product incorporating part of the 3' LTR. PCR products can be cloned and sequenced. Adapted from Pearce (1999).

B.2.3.PCR conditions

50ng of purified template DNA was PCR amplified in 25µl of reaction buffer consisting of 1X PCR buffer (Qiagen), 0.8µg of RNaseH motif 1 primer (5'-MGNACNAARCAYATHGA-3'), 0.15µg of MseI adapter primer (5'-GATGGATCCTGAGTAA-3'), 200µM DNTPs and 1 unit of hotstar taq DNA

polymerase (Qiagen). PCR reactions were incubated at 95ºC for 15min followed by 30 cycles of: 94ºC for 60sec; 45ºC for 90sec; 72ºC for 90sec, then 72ºC for 7min. 5µl of the PCR product was visualised by agarose gel electrophoresis. Successful reactions showed a smear of PCR products with a mean size of 500bp. A previous protocol loaded all of the PCR reaction onto an agarose gel, cut out all products in the size range 300-600bp before recovering the PCR products using a 'mini-elute' gel extraction kit (Qiagen). This method was discontinued when it was found that it made no difference to the size range of PCR products produced in the 2nd PCR.

2µl of PCR products were PCR amplified in 25µl of reaction buffer consisting of 1X PCR buffer (Qiagen), 0.8µg of RNaseH motif 2 primer (5'-GCNGAYATNYTNACNAA-3'), 0.15µg of MseI adapter primer (5'-GATGGATCCTGAGTAA-3'), 200µM DNTPs and 1 unit of hotstar taq DNA polymerase (Qiagen). Reaction conditions were identical to those described for the 1st PCR reaction. Again, 5µl of PCR product was visualised by agarose gel electrophoresis. The PCR products were purified using a qiaquick PCR purification column according to the manufacturer's instructions.

B.2.4.Cloning

LB growth medium was prepared by dissolving 5g of NaCl, 2.5g of yeast extract and 5g of tryptone in 1 litre of distilled water. The pH was adjusted to 7 using HCl or NaOH. LB was divided into 250ml flasks, for solid growth medium, 3.5g of bacteriological agar. Flasks of LB growth medium were sterilised in an autoclave. Agar plates were prepared by melting 250ml of solid growth medium; flasks were allowed to cool on a rotary shaker, before the growth medium set, 800µl of X-gal (20mg/ml in formamide) and 100 µl of ampicillin (100µg/ml). The growth medium was poured into sterile petri plates in a flow hood.

PCR products were ligated into pGEM-T-easy vectors (Promega). 3µl of purified PRC product was placed in a 0.2 ml PCR tube along with 5µl of 2x reaction buffer, 1µl of T4 DNA ligase (3u/µl) and 1ml of pGEM-T-easy vector (50ng/µl). Ligation reactions were incubated at 4ºC for 16h.

Vectors containing ligated PCR products were transformed into cells and cultured on agar plates. JM109 high efficiency competent cells were gently thawed on ice (approximately 5 min). 50 µl of cells were pippetted into 1.5 microcentrifuge tubes along with 7.5µl of the ligation reaction. Transformation reactions were placed on ice for 20min, heat shocked at 42ºC for 90s and immediately returned to ice for 2min. 500µl of LB growth medium was added, and the reaction incubated at 37ºC on a rotary shaker for 1h. Transformed cells were plated on three agar plates; one with 50µl of the transformation reaction, another with 200µl, and the final plate with the remainder. The agar plates were incubated at 37ºC for 16h.

B.2.5.Determination of insertion size

White or pale blue colonies were picked from the agar plates, using a sterile toothpick. The selected colonies were plated in duplicate on a separate agar plate, and transferred to a 0.2µl PCR tube containing 10µl of 1X PCR reaction buffer (Bioline) containing 0.4µg of RNaseH motif 2 primer, 0.075µg of MseI adapter primer, 200mM DNTPs, and 0.5 units of Redtaq (Bioline). PCR conditions were identical to those described for the initial PCR reactions. Reactions were visualised by agarose gel electrophoresis. Colonies with inserts larger than approximately 300bp were selected for sequencing.

B.2.6.Isolation of plasmid DNA

Selected colonies were removed from the plate using a sterile toothpick and transferred to a 30ml universal glass tube containing 5ml of LB containing 2.5µl ampicillin (100mg/ml). Cultures were incubated at 37ºC for 16h on a rotary shaker. 1.25ml of the culture was centrifuged at 13 000 rpm for 20s. Another 1.25ml was transferred to the same tube, and the centrifuge step repeated. Plasmid extractions were conducted using a 'Perfectprep mini kit' (Eppendorf) according to the manufacturer's instructions. Extracted plasmid DNA was eluted in 50µl of elution buffer. Plasmid concentration was determined on a 1% agarose gel using 25, 50 and 100ng λ DNA.

B.2.7.Cycle sequencing

50ng plasmid DNA was placed in the well of a 96 well plate and the volume made up to 10µl using mili-Q water. The plates were covered with a silicone lid. Plasmid DNA was heated to 96ºC for 1 minute before being returned to room temperature. This stage ensured that plasmid DNA was uncoiled for the sequencing reactions. 8µl of DTCS was added to each reaction well along with 2µl of the appropriate primer (Forward: 5′-CGCCAGGGTTTTCCCAGTCAG-3′, Reverse: 5′-CACAGGAAACAGCTATGAC-3′). Thermal cycling consisted of 30 cycles of: 96ºC for 20s, 50ºC for 20s, and 60ºC for 4 minutes. After thermal cycling, reactions were held at 4ºC.

Sequencing reactions were purified by ethanol precipitation. 5µl of stop solution (2µl 3M NaOAc; 2µl 100mM EDTA; 1µl glycogen) was added to the reaction. 60µl of 95% ethanol was added to each sample and mixed thoroughly. The sample was centrifuged at 13 000 rpm at 4ºC for 5 min to pellet the DNA. The supernatant was carefully removed and 200µl of 70% ethanol added. Samples were again centrifuged for 5 min at 4ºC and the supernatant removed. 70% ethanol was again added, and the process repeated. Samples were placed in a vacuum chamber and dried for 20min. Dried pellets were re-suspended in 40µl of sample loading solution (Beckman Coulter). Each sample was transferred to a well of a 96 well plate and overlaid with mineral oil.

Sequence reactions were processed using a Beckman Coulter CEQ 8000 automated sequencer. Sequence analysis was conducted using the sequence analysis module of the CEQ software. Analysed sequence results were exported to genedoc (Nicholas and Nicholas, 1997) and aligned. Sequence alignments were manually checked using the investigator module of the CEQ software.

B.2.8.Identification of LTR sequence

Following RNaseH motif 2, the RNaseH open reading frame continues for a variable number of base pairs (Pearce et al., 1999). As such, it is not possible to identify the RNaseH/ LTR junction sequence bases solely on its position within the sequence.

Instead, (T)TG or (T)TGT sequence motifs following a polypurine tract, which are characteristic of the 5'end of a LTR, were identified as putative markers of LTR junction sequence.

B.2.9.Design of LTR primers

Primers were designed from the LTR junction sequence to amplify outwards from the 5′ LTR. Initially primers were located immediately adjacent to this junction and with a selective base in order that they would not amplify from the 3′. However, this approach was abandoned in order to allow more control over primer design during later attempts to isolate 5′ flanking sequence. All primers were designed using the AmplifX computer program (Jullien, 2006).

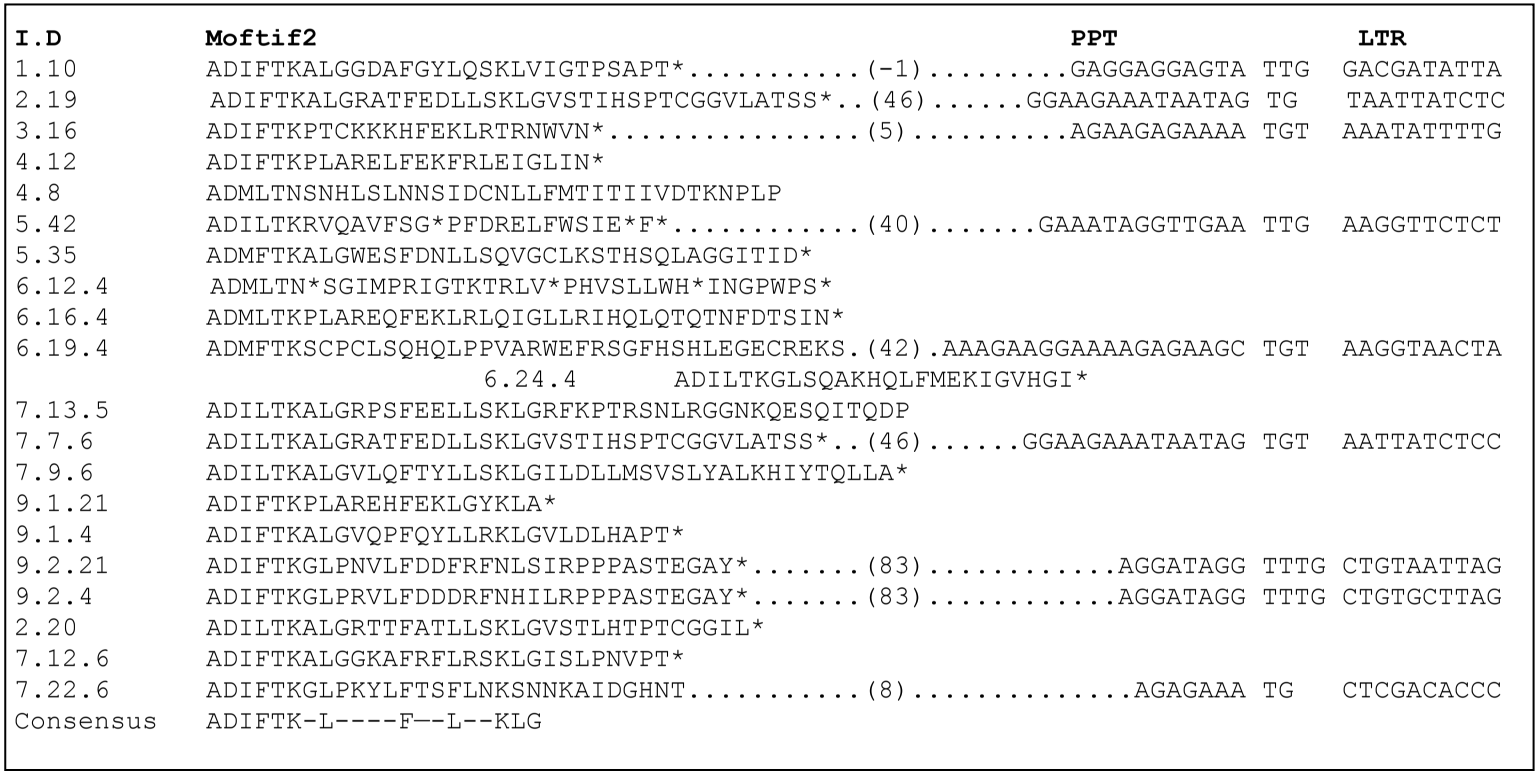
B.3. Results

B.3.1.LTR sequencing

40 plasmids were extracted for sequencing. Of these, 24 produced sequence data of sufficient quality. Of these, 9 sequences showed evidence for a putative PPT/ LTR junction. These are illustrated in Figure B.4.

B.3.2.Isolation of flanking sequence

Of the nine putative LTR sequences identified, primers were designed from the three most likely candidates. These were: 1.10, 5.42, and 19.4. Initially, primers were designed at the PPT/ LTR junction; this did not allow the specificity of the sequences to be tested. Sequences derived from reactions using these primers were used to design primers to isolate 3′ sequence. However, when these primers were used in accessions from which LTR sequence was derived, they failed to regenerate LTR sequence. This suggested that the LTR primers were binding to non-specific sites. This was confirmed when these primers were redesigned for a site within the LTR. In all cases, the sequences generated from resulting PCR reactions failed to show any homology with LTR sequence from which the primers were designed outside of the primer binding site. Despite extensive optimisation of the PCR conditions, none of the primers designed from LTR sequence generated target specific sequences.



I.D	Moftif2	PPT	LTR
1.10	ADIFTKALGGDAFGYLQSKLVIGTPSAPT*.....(-1).....	GAGGAGGAGTA TTG	GACGATATTA
2.19	ADIFTKALGRATFEDLLSKLGVSTIHSPTCGGVLATSS*..(46).....	GGAAGAAATAATAG TG	TAATTATCTC
3.16	ADIFTKPTCKKKHFEKLRTRNWVN*.....(5).....	AGAAGAGAAAA TGT	AAATATTTTG
4.12	ADIFTKPLARELFEKFRLEIGLIN*		
4.8	ADMLTNSNHLSLNNSIDCNLLFMTITIIVDTKNPLP		
5.42	ADILTKRVQAVFSG*PFDRELFWSIE*F*.....(40).....	GAAATAGGTTGAA TTG	AAGGTTCTCT
5.35	ADMFTKALGWESFDNLLSQVGCLKSTHSQLAGGITID*		
6.12.4	ADMLTN*SGIMPRIGTKTRLV*PHVSLLWH*INGPWPS*		
6.16.4	ADMLTKPLAREQFEKLRLQIGLLRIHQLQTQTNFDTSIN*		
6.19.4	ADMFTKSCPCLSQHQLPPVARWEFRSGFHSHLEGECREKS.(42).	AAAGAAGGAAAAGAGAAGC TGT	AAGGTAACTA
6.24.4	ADILTKGLSQAKHQLFMEKIGVHGI*		
7.13.5	ADILTKALGRPSFEELLSKLGRFKPTRSNLRGGNKQESQITQDP		
7.7.6	ADILTKALGRATFEDLLSKLGVSTIHSPTCGGVLATSS*..(46).....	GGAAGAAATAATAG TGT	AATTATCTCC
7.9.6	ADILTKALGVLQFTYLLSKLGILDLLMSVSLYALKHIYTQLLA*		
9.1.21	ADIFTKPLAREHFEKLGYKLA*		
9.1.4	ADIFTKALGVQPFQYLLRKLGVLDLHAPT*		
9.2.21	ADIFTKGLPNVLFDDFRFNLSIRPPPASTEGAY*.....(83).....	AGGATAGG TTTG	CTGTAATTAG
9.2.4	ADIFTKGLPRVLFDDDRFNHILRPPPASTEGAY*.....(83).....	AGGATAGG TTTG	CTGTGCTTAG
2.20	ADILTKALGRTTFATLLSKLGVSTLHTPTCGGIL*		
7.12.6	ADIFTKALGGKAFRFLRSKLGISLPNVPT*		
7.22.6	ADIFTKGLPKYLFTSFLNKSNNKAIDGHNT.....(8).....	AGAGAAA TG	CTCGACACCC
Consensus	ADIFTK-L----F--L--KLG		

Figure B.4: Amino acid/ nucleotide sequence alignment , showing the RNaseH motif 2, Popyypurine tract (PPT) and LTR region. Where intervening sequence is not shown, numbers in parentheses indicate the number of intervening nucleotide bases.

B.4. Conclusions

The failure to develop co-dominant molecular markers from LTR retrotransposons is disappointing. It seems likely that this is due, at least in part, to the constraints imposed on the selection of the primer binding site, particularly when primer sites were located exactly on the PPT/ LTR junction. Easing this restriction, and designing primers internal to the LTR would be expected to improve results but even this proved ineffective.

The quality and specificity of primers that are designed to amplify flanking sequence from this type of retrotransposons would seem to be a major limitation to the development of co-dominant markers from retrotransposons. This suggests that a greater sequencing effort in order to generate a large number of LTR sequences, in which suitable primer sites might be found would seem to be the most effective strategy, both for the development of SSAP or RBIP type markers.

References

ANDOLFATTO, P. 2005. Adaptive evolution of non-coding DNA in Drosophila. *Nature* 437: 1149-1152.

ARABIDOPSIS-GENOME-INITIATIVE. 2002. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815.

ARUMUGANATHAN, K., AND E. D. EARLE. 1991. Estimation of nuclear DNA content of plants by flow cytometry. *Plant Molecular Biology Reporter* 9: 217-229.

BARRE, P., M. LAYSSAC, A. D'HONT, J. LOUARN, A. CHARRIER, S. HAMON, AND M. NOIROT. 1998. Relationship between parental chromosomic contribution and nuclear DNA content in the coffee interspecific hybrid *C. pseudozanguebariae* X *C. liberica var. 'dewevre*i'. *Theoretical and Applied Genetics* 96: 301-305.

BATEMAN, A. J. 1948. Intrasexual selection in *Drosophila*. *Heredity* 2: 349-368.

BEADLE, G. W., AND E. L. TATUM. 1941. Genetic control of biochemical reactions in *Neurospora*. *Proceedings of the National Academy of Sciences of the USA* 27: 499-506.

BELL, G. 1985. On the Function of Flowers. *Proceedings of the Royal Society of London Series B-Biological Sciences* 224: 223-265.

BENNETT, M. D. 1971. The duration of meiosis. *Proceedings of the Royal Society of London Series B-Biological Sciences* 178: 277-299.

BENNETT, M. D. 1972. Nuclear DNA content and minimum generation time in herbaceous plants. *Proceedings of the Royal Society of London Series B-Biological Sciences* 181: 109-135.

BENNETT, M. D. 1976. DNA Amount, Latitude, and Crop Plant Distribution. *Environmental and Experimental Botany* 16: 93-108.

BENNETT, M. D., AND J. B. SMITH. 1976. Nuclear DNA amounts in angiosperms. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 274: 227-274.

BENNETT, M. D., AND I. J. LEITCH. 1995. Nuclear DNA amounts in angiosperms. *Annals of Botany* 76: 113-176.

BENNETT, M. D., AND I. J. LEITCH. 2004. http://www.rbgkew.org.uk/cval/homepage.html.

BENNETT, M. D., J. B. SMITH, AND J. S. HESLOP-HARRISON. 1982. Nuclear DNA amounts in angiosperms. *Proceedings of the Royal Society of London Series B-Biological Sciences* 216: 179-199.

BENNETT, M. D., P. BHANDOL, AND I. J. LEITCH. 2000. Nuclear DNA amounts in angiosperms and their modern uses: 807 new estimates. *Annals of Botany* 85: 351-357.

BENNETZEN, J. L. 1996. The contributions of retroelements to plant genome organization, function and evolution. *Trends Microbiol* 4: 347-353.

BENNETZEN, J. L. 2000. Transposable element contributions to plant gene and genome evolution. *Plant Molecular Biology* 42: 251-269.

BENNETZEN, J. L., AND E. A. KELLOGG. 1997. Do plants have a one-way ticket to genomic obesity? *Plant Cell* 9: 1509-1514.

BIDERRE, C., M. PAGES, G. METENIER, E. U. CANNING, AND C. P. VIVARES. 1995. Evidence for the smallest nuclear genome (2.9Mbp) in the microsporidiun

Encephalitozoon cuniculi. *Molecular and Biochemical Parisitology* 74: 229-233.

BINGHAM, P. M., AND Z. ZACHAR. 1989. Retrotransosons and FB elements from *Drosophila melanogaster*. *In* D. E. Berg and M. M. Howe [eds.], Mobile DNA, 485-502. American Society of Microbiology, Washington DC.

BOEKE, J. D., AND V. G. CORRES. 1989. Transcription and reverse transcription of retrotransposons. *Annual Review of Microbiology* 43: 403-434.

BRANTJES, N. B. 1976. Riddles around the pollination of *Melandrium album* (Mill.) Garcke (Carophyllacea) during the oviposition by *Hadena bicruris Hufn*. (Noctiudae, Lepidoptera). *Proceedings Koninkl. Nederlandse Akademie Wetenschappen. Amsterdam, Series C* 79: 1-12.

BRITTEN, R. J., AND D. E. KOHNE. 1968. Repeated sequences in DNA. *Science* 161: 529-540.

BURES, P., Y. WANG, L. HOROVA, AND J. SUDA. 2004. Genome size variation in central European species of *Cirsium* (Compositae) and their natural hybrids. *Annals of Botany* 94: 353-363.

CACERES, M. E., C. DE PACE, G. T. SCARASCIA MUGNOZZA, P. KOTSONIS, M. CECCARELLI, AND P. G. CIONINI. 1998. Genome size variations within *Dasypyrum villosum* : correlations with chromosomal traits, environmental factors and plant phenotypic characteristics and behaviour in reproduction. *Theoretical and Applied Genetics* 96: 559-567.

CAETANO-ANOLLES, G. 2005. Evolution of genome size in the grasses. *Crop Science* 45: 1809-1816.

CAVALIER-SMITH, T. 1978. Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate and the solution of the DNA C-value paradox. *Journal of Cell Science* 34: 247-278.

CAVALIER-SMITH, T. 1982. Skeletal DNA and the evolution of genome size. *Annual Review of Biophysics and Bioengineering* 11: 273-302.

CAVALIER-SMITH, T. 1985. Introduction: the evolutionary significance of genome size. *In* T. Cavalier-Smith [ed.], The Evolution of Genome Size, 1-36. John Wiley and Sons, Chichester.

CAVALIER-SMITH, T., AND M. J. BEATON. 1999. The skeletal function of non-genic nuclear DNA: new evidence from ancient cell chimaeras. *Genetica* 106: 3-13.

CHARLESWORTH, B. 1986. Genetic-Divergence between Transposable Elements. *Genetical Research* 48: 111-118.

CHARLESWORTH, B., P. SNIEGOWSKI, AND W. STEPHAN. 1994. The Evolutionary Dynamics of Repetitive DNA in Eukaryotes. *Nature* 371: 215-220.

CHARLESWORTH, D. 2002. Plant sex determination and sex chromsomes. *Heredity* 88: 94-101.

CHUNG, J., J. H. LEE, K. ARUMUGANATHAN, G. L. GRAEF, AND J. E. SPECHT. 1998. Relationships between nuclear DNA content and seed and leaf size in soybean. *Theoretical and Applied Genetics* 96: 1064-1068.

CIUPERCESCU, D. D., J. VEUSKENS, A. MOURAS, D. YE, M. BRIQUET, AND I. NEGRUTIU. 1990. Karyotyping *Melandrium album*, a dioecious plant with heteromorphic sex-chromosomes. *Genome* 33: 556-562.

COMMONER, B. 1964. Roles of deoxyribonucleic acid in inheritance. *Nature* 202: 960-968.

COSTICH, D. E., T. R. MEAGHER, AND E. J. YURKOW. 1991. A rapid means of sex identification in *Silene latifolia* by use of flow cytometry. *Plant Molecular Biology* 9: 359-370.

COYNE, J. A. 1985. The genetic-basis of Haldane's rule. *Nature* 314: 736-738.
DEELY, E. M. 1955. An integrating microdensitometer for biological cells. *Journal of Cell Instrumentation* 32: 263-267.
DELPH, L. F., L. F. GALLOWAY, AND M. L. STANTON. 1996. Sexual dimorphism in flower size. *American Naturalist* 148: 299-320.
DELPH, L. F., F. N. KNAPCZYK, AND D. R. TAYLOR. 2002. Among-population variation and correlations in sexually dimorphic traits of *Silene latifolia*. *Journal of Evolutionary Biology* 15: 1011-1020.
DELPH, L. F., F. M. FREY, J. C. STEVEN, AND J. L. GEHRING. 2004. Investigating the independent evolution of the size of floral organs via G-matrix estimation and artificial selection. *Evolution and Development* 6: 438-448.
DOLEZEL, J., AND W. GOHDE. 1995. Sex determination in dioecious plants *Melandrium album* and *M. rubrum* using high resolution flow-cytometry. *Cytometry* 19: 103-106.
DOLEZEL, J., AND J. BARTOS. 2005. Plant DNA flow cytometry and estimation of nuclear genome size. *Annals of Botany* 95: 99-110.
DOOLITTLE, W. F., AND C. SAPIENZA. 1980. Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284: 601-603.
EVANS, G. M., H. REES, C. L. SNELL, AND S. SUN. 1973. The relationship between nuclear DNA amount and the duration of the mitotic cycle. *Chromosomes Today* 3: 24-31.
FALSTER, D. S., D. I. WARTON, AND I. J. WRIGHT. 2006. SMATR: Standardised major axis tests and routines, ver 2.0.
FAY, M. F., R. S. CONWAN, AND I. J. LEITCH. 2005. The effects of nuclear DNA content (C-value) on the quality and utility of AFLP fingerprints. *Annals of Botany* 95: 237-246.
FILATOV, D. A., AND D. CHARLESWORTH. 2002. Substitution rates in the X- and Y-linked genes of the plants, *Silene latifolia* and *S. dioica*. *Molecular Biology and Evolution* 19: 898-907.
FLAVELL, A. J., D. B. SMITH, AND A. KUMAR. 1992a. Extreme heterogeneity of Ty1-*copia* group retrotransposons in plants. *Molecular & General Genetics* 231: 233-242.
FLAVELL, A. J., M. D. BENNETT, J. B. SMITH, AND D. B. SMITH. 1974. Genome size and the proportion of repeated nucleotide sequence DNA in plants. *Biochemical Genetics* 12: 257-259.
FLAVELL, A. J., S. R. PEARCE, J. S. P. HESLOP-HARRISON, AND A. KUMAR. 1997. The evolution of Ty1-*copia* group retrotransposons in eukaryote genomes. *Genetica* 100: 185-195.
FLAVELL, A. J., M. R. KNOX, S. R. PEARCE, AND T. H. N. ELLIS. 1998. Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. *Plant Journal* 16: 643-650.
FLAVELL, A. J., E. DUNBAR, R. ANDERSON, S. R. PEARCE, R. HARTLEY, AND A. KUMAR. 1992b. Ty1-*copia* group retrotransposons are ubiquitous and heterogeneous in higher-plants. *Nucleic Acids Research* 20: 3639-3644.
GALLOWAY, L. F., AND J. R. ETTERSON. 2005. Population differentiation and hybrid success in *Campanula americana*: geography and genome size. *Journal of Evolutionary Biology* 18: 81-89.
GARNATJE, T., J. VALLES, S. GARCIA, O. HIDALGO, M. SANZ, M. A. CANELA, AND S. SILJAK-YAKOVLEV. 2004. Genome size in *Echinops L*. and related genera

(Asteraceae, Cardueae): karyological, ecological and phylogenetic implications. *Biology of the Cell* 96: 117-124.

GOULSON, D., AND K. JERRIM. 1997. Maintenance of the species boundary between *Silene dioica* and *S-latifolia* (red and white campion). *Oikos* 79: 115-126.

GRABOWSKA-JOACHIMIAK, A., AND A. JOACHIMIAK. 2002. C-banded karyotypes of two *Silene* species with heteromorphic sex chromosomes. *Genome* 45: 243-252.

GRANT, S., A. HOUBEN, B. VYSKOT, J. SIROKY, W. H. PAN, J. MACAS, AND H. SAEDLER. 1994. Genetics of sex determination in flowering plants. *Developmental Genetics* 15: 214-230.

GREGORY, T. R. 2001. Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biological Reviews* 76: 65-101.

GREGORY, T. R., P. D. HEBERT, AND J. KOLASA. 2000. Evolutionary implications of the relationship between genome size and body size in flatworms and copepods. *Heredity* 84 (Pt 2): 201-208.

GREILHUBER, J. 1979. Evolutionary changes of DNA and heterochromatin amounts in the *Scilla bifolia* group (Liliaceae). *Plant Systematics and Evolution, Supplement* 2: 263-280.

GREILHUBER, J. 1986. Severely distorted feulgen DNA amounts in *Pinus* (Coniferophytina) after nonadditive fixations as a result of meristematic self-tanning with vacuole contents. *Canadian Journal of Genetics and Cytology* 28: 409-415.

GREILHUBER, J. 1988a. Critical reassessment of DNA content variation in plants. *In* P. E. Brandham [ed.], Kew chromosome conference III, 39-50. HMSO, London.

GREILHUBER, J. 1988b. 'Self-tanning'—a new and important source of stoichiometric error in cytophotometric determination of nuclear DNA content in plants. *Plant Systematics and Evolution* 158: 87-96.

GREILHUBER, J. 1997. The problem of variable genome size in plants (with special reference to woody plants). *In* Z. Borzan and S. E. Schlarbaum [eds.], Cytogenetic studies of forest trees and shrub species. Proceedings of the first IUFRO Cytogenetics Working Party, Brijunji National Park, Croatia, September 8-11, 1993., 13-34. Croatian Forests Inc., Faculty of Forestry, University of Zagreb, Zagreb.

GREILHUBER, J. 1998. Intraspecific variation in genome size: A critical reassessment. *Annals of Botany* 82: 27-35.

GREILHUBER, J. 2005. Intraspecific variation in genome size in angiosperms: Identifiving its existence. *Annals of Botany* 95: 91-98.

GREILHUBER, J., AND F. SEPTA. 1985. Geographical variation of genome size at low taxanomic levels in the *Scilla bifolia* alliance (Hyacinthaceae). *Flora* 176: 431-438.

GREILHUBER, J., AND R. OBERMAYER. 1997. Genome size and maturity group in *Glycine max* (soybean). *Heredity* 78: 547-551.

GRIME, J. P., AND M. A. MOWFORTH. 1982. Retrotransposon-mediated genome evolution on a local ecological scale. *Nature* 299: 151-153.

GROSSET, L., AND N. ODARTCHENKO. 1975. Duration of mitosis and separate mitotic phases compared to nuclear DNA content in erythroblasts of four vertebrates. *Cell and Tissue Kinetics* 8: 91-96.

HALDANE, J. B. S. 1922. Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics* 12: 101-109.

HALL, S. E., W. S. DVORAK, J. S. JOHNSTON, H. J. PRICE, AND C. G. WILLIAMS. 2000. Flow cytometric analysis of DNA content for tropical and temperate new world pines. *Annals of Botany* 86: 1081-1086.

HARDIE, D. C., AND P. D. N. HEBERT. 2003. The nucleotypic effects of cellular DNA content in cartilaginous and ray-finned fishes. *Genome* 46: 683-706.

HAUBOLD, B., AND T. WIEHE. 2006. How repetitive are genomes? *Bmc Bioinformatics* 7: 541.

HIROCHIKA, H., AND R. HIROCHIKA. 1993. Ty1-*copia* group retrotransposons as ubiquitous components of plant genomes. *Japanese Journal of Genetics* 68: 35-46.

HONDA, M., Y. MURAMOTO, T. KUZUGUCHI, S. SAWANO, M. MACHIDA, AND H. KOYAMA. 2002. Determination of gene copy number and genotype of transgenic *Arabidopsis thaliana* by competitive PCR. *Journal of Experimental Botany* 53: 1515-1520.

HUELSENBECK, J. P., AND F. RONQUIST. 2001. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 17: 754-755.

HUMEAU, L., T. PAILLER, AND J. D. THOMPSON. 2003. Flower size dimorphism in diclinous plants native to La Reunion Island. *Plant Systematics and Evolution* 240: 163-173.

IRICK, H. 1994. A new function for heterochromatin. *Chromosoma* 103: 1-3.

JOHNSTON, J. S., M. D. BENNETT, A. L. RAYBURN, D. W. GALBRAITH, AND H. J. PRICE. 1999. Reference standards for determination of DNA content of plant nuclei. *American Journal of Botany* 86: 609.

JOHNSTON, J. S., A. E. PEPPER, A. E. HALL, Z. J. CHEN, G. HODNETT, J. DRABEK, R. LOPEZ, AND H. J. PRICE. 2005. Evolution of genome size in Brassicaceae. *Annals of Botany* 95: 229-235.

JULLIEN, N. 2006. AmplifX.

JURGENS, A., T. WITT, AND G. GOTTSBERGER. 1996. Reproduction and pollination in central European populations of *Silene* and *Saponaria* species. *Botanica Acta* 109: 316-324.

KALENDAR, R., J. TANSKANEN, S. IMMONEN, E. NEVO, AND A. H. SCHULMAN. 2000. Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE-1* retrotransposon dynamics in response to sharp microclimatic divergence. *Proceedings of the National Academy of Sciences of the USA* 97: 6603-6607.

KATO, H., P. SRIPRASERTSAK, H. SEKI, Y. ICHINOSE, T. SHIRAISHI, AND T. YAMANDA. 1999. Functional analysis of retrotransposons in pea. *Plant and Cell Physiology* 40: 933-941.

KEJNOVSKY, E., Z. KUBAT, R. HOBZA, M. LENGEROVA, S. SATO, S. TABATA, K. FUKUI, S. MATSUNAGA, AND B. VYSKOT. 2006. Accumulation of chloroplast DNA sequences on the Y chromosome of *Silene latifolia*. *Genetica* 128: 167-175.

KELLOGG, E. A., AND J. L. BENNETZEN. 2004. The evolution of nuclear genome structure in seed plants. *American Journal of Botany* 91: 1709-1725.

KIDWELL, M. G. 2002. Transposable elements and the evolution of genome size in eukaryotes. *Genetica* 115: 49-63.

KIMURA, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

KIRIK, A., S. SALOMON, AND H. PUCHTA. 2000. Species-specific double-strand break repair and genome evolution in plants. *Embo Journal* 19: 5562-5566.

KNIGHT, C. A., N. A. MOLINARI, AND D. A. PETROV. 2005. The large genome constraint hypothesis: Evolution, ecology and phenotype. *Annals of Botany* 95: 177-190.

KUBIS, S., T. SCHMIDT, AND J. S. HESLOP-HARRISON. 1998. Repetitive DNA elements as a major component of plant genomes. *Annals of Botany* 82: 45-55.

KULIKOVA, T., R. AKHTAR, P. ALDEBERT, N. ALTHORPE, M. ANDERSSON, A. BALDWIN, K. BATES, S. BHATTACHARYYA, L. BOWER, P. BROWNE, M. CASTRO, G. COCHRANE, K. DUGGAN, R. EBERHARDT, N. FARUQUE, G. HOAD, C. KANZ, C. LEE, R. LEINONEN, Q. LIN, V. LOMBARD, R. LOPEZ, D. LORENC, H. MCWILLIAM, G. MUKHERJEE, F. NARDONE, M. PILAR, G. PASTOR, S. PLAISTER, S. SOBHANY, P. STOEHR, R. VAUGHAN, D. WU, W. M. ZHU, AND R. APWEILER. 2007. EMBL Nucleotide Sequence Database in 2006. *Nucleic Acids Research* 35: D16-D20.

KUMAR, A. 1996. The adventures of the Ty1-*copia* group of retrotransposons in plants. *Trends in Genetics* 12: 41-43.

KUMAR, A., AND J. L. BENNETZEN. 1999. Plant retrotransposons. *Annual Review of Genetics* 33: 479-532.

KUMAR, A., AND H. HIROCHIKA. 2001. Applications of retrotransposons as genetic tools in plant biology. *Trends in Plant Science* 6: 127-133.

L'HOMME, Y., A. SEGUIN, AND F. M. TREMBLAY. 2000. Different classes of retrotransposons in coniferous spruce species. *Genome* 43: 1084-1089.

LAGERCRANTZ, U., H. ELLEGREN, AND L. ANDERSSON. 1993. The Abundance of Various Polymorphic Microsatellite Motifs Differs between Plants and Vertebrates. *Nucleic Acids Research* 21: 1111-1115.

LASSNER, M. W., P. PETTERSON, AND J. I. YODER. 1989. Simultaneous amplification of multiple DNA fragments by polymerase chain reaction in the analysis of transgenic plants and their progeny. *Plant Molecular Biology Reporter* 7: 116-128.

LAURIE, D. A., AND M. D. BENNETT. 1985. Nuclear-DNA Content in the Genera *Zea* and *Sorghum* - Intergeneric, Interspecific and Intraspecific Variation. *Heredity* 55: 307-313.

LAWRENCE, C. W. 1963. Genetic studies on wild populations of *Melandrium*. II. Flowering time and plant weight. *Heredity* 18: 149-163.

LEITCH, I. J., AND M. D. BENNETT. 1997. Polyploidy in angiosperms. *Trends in Plant Science* 2: 470-476.

LENGEROVA, M., R. C. MOORE, S. R. GRANT, AND B. VYSKOT. 2003. The sex chromosomes of *Silene latifolia* revisited and revised. *Genetics* 165: 935-938.

LEVIN, D. A., AND S. W. FUNDERBURG. 1979. Genome Size in Angiosperms - Temperate Versus Tropical Species. *American Naturalist* 114: 784-795.

LLOYD, D. G., AND C. J. WEBB. 1977. Secondary sex characters in plants. *Botanical Review* 43: 177-216.

LYONS, E. E., D. MILLER, AND T. R. MEAGHER. 1994. Evolutionary Dynamics of Sex-Ratio and Gender Dimorphism in *Silene Latifolia* .1. Environmental-Effects. *Journal of Heredity* 85: 196-203.

LYONS, E. E., N. SHAHMAHONEY, AND L. A. LOMBARD. 1995. Evolutionary Dynamics of Sex-Ratio and Gender Dimorphism in *Silene Latifolia* .2. Sex-Ratio and Flowering Status in a Potentially Male-Biased Population. *Journal of Heredity* 86: 107-113.

MATSUNAGA, S., F. YAGISAWA, M. YAMAMOTO, W. UCHIDA, S. NAKAO, AND S. KAWANO. 2002. LTR retrotransposons in the dioecious plant *Silene latifolia*. *Genome* 45: 745-751.

MATSUOKA, Y., AND K. TSUNEWAKI. 1996. Wheat retrotransposon families identified by reverse transcriptase domain analysis. *Molecular Biology and Evolution* 13: 1384-1392.

MATSUOKA, Y., AND K. TSUNEWAKI. 1999. Evolutionary dynamics of Ty1-*copia* group retrotransposons in grass shown by reverse transcriptase domain analysis. *Molecular Biology and Evolution* 16: 208-217.

MEAGHER, T. R. 1992. The Quantitative Genetics of Sexual Dimorphism in *Silene latifolia* (Caryophyllaceae) .1. Genetic-Variation. *Evolution* 46: 445-457.

MEAGHER, T. R. 1994. The quantitative genetics of sexual dimorphism in *Silene latifolia* (Caryophllaceae).II. Response to sex -specific selection. *Evolution* 48: 939-951.

MEAGHER, T. R. 1999. Quanitative genetics of sexual dimorphism. *In* M. A. Geber, T. E. Dawson, and L. F. Delph [eds.], Gender and sexual dimorphism in flowering plants., 275-294. Springer-Verlag, New York.

MEAGHER, T. R., AND D. E. COSTICH. 1994. Sexual Dimorphism in Nuclear-DNA Content and Floral Morphology in Populations of *Silene latifolia* (Caryophyllaceae). *American Journal of Botany* 81: 1198-1204.

MEAGHER, T. R., AND D. E. COSTICH. 1996. Nuclear DNA content and floral evolution in *Silene latifolia*. *Proceedings of the Royal Society of London Series B-Biological Sciences* 263: 1455-1460.

MEAGHER, T. R., AND D. E. COSTICH. 2004. 'Junk' DNA and long-term phenotypic evolution in *Silene* section *Elisanthe* (Caryophyllaceae). *Proceedings of the Royal Society of London Series B-Biological Sciences* 271: S493-S497.

MEAGHER, T. R., AND C. VASSILIADIS. 2005. Phenotypic impacts of repetitive DNA in flowering plants. *New Phytologist* 168: 71-80.

MILLER, J. S., AND D. L. VENABLE. 2003. Floral morphometrics and the evolution of sexual dimorphism in *Lycium* (Solanaceae). *Evolution* 57: 74-86.

MILLER, W. G. 2007. Openstat.

MINELLI, S., P. MOSCARIELLO, AND P. G. CIONINI. 1996. Nucleotype and phenotype in *Vicia faba*. *Heredity* 76: 524-530.

MINITAB-LTD. 2006. Minitab 15.

MIRSKY, A. E., AND H. RIS. 1951. The deoxyribonucleic acid content of animal cells and its evolutionary significance. *Journal of General Physiology* 34: 451-462.

MORGAN, E. R., G. K. BURGE, J. F. SEELYE, M. E. HOPPING, AND J. E. GRANT. 1998. Production of interspecific hybrids *Limonium perezii* (Stapf) Hubb. and *Limonium sinuatum* (L.) Mill. *Euphytica* 102: 109-115.

MOSCONE, E. A., M. BARYANI, I. EBERT, J. GREILHUBER, F. EHRENDORFER, AND A. T. HUNZIKER. 2003. Analysis of nuclear DNA content in *Capsicum* (Solanaceae) by flow cytometry and Feulgen densitometry. *Annals of Botany* 92: 21-29.

MOUNT, S. M., AND G. M. RUBIN. 1985. Complete nucleotide sequence of the *Drosophila* transposable element *copia*: homology between *copia* and retroviral protiens. *Molecular and Cellular Biology* 5: 1630-1638.

MOWFORTH, M. A., AND J. P. GRIME. 1989. Intra-population variation in nuclear DNA amount, cell size and growth rate in *Poa annua* L. *Functional Ecology* 3: 289-295.

MURRAY, B. G. 2005. When does intraspecific C-value variation become taxonomically significant? *Annals of Botany* 95: 119-125.

NAGL, W. 1974. Mitotic cycle time in perennial and annual plants with various amounts of DNA and heterochromatin. *Developmental Biology* 39: 342-346.

NAVARRO-QUEZADA, A., AND D. J. SCHOEN. 2002. Sequence evolution and copy number of Ty1-*copia* retrotransposons in diverse plant genomes. *Proceedings of the National Academy of Sciences of the United States of America* 99: 268-273.

NICHOLAS, K. B., AND H. B. NICHOLAS. 1997. Genedoc: a tool for editing and annotating multiple sequence alignments.

NOIROT, M., P. BARRE, J. LOUARN, C. DUPERRAY, AND S. HAMON. 2000. Nucleus-cytosol interactions—a source of stoichiometric error in flow cytometric estimation of nuclear DNA content in plants. *Annals of Botany* 86: 309-316.

NOIROT, M., P. BARRE, J. LOUARN, C. DUPERRAY, AND S. HAMON. 2002. Consequences of stoichiometric error on nuclear DNA content evaluation in *Coffea liberica var. dewevrei* using DAPI and propidium iodide. *Annals of Botany* 89: 385-389.

NOIROT, M., P. BARRE, C. DUPERRAY, J. LOUARN, AND S. HAMON. 2003. Effects of caffeine and chlorogenic acid on propidium iodide accessibility to DNA: Consequences on genome size evaluation in coffee tree. *Annals of Botany* 92: 259-264.

NOIROT, M., P. BARRE, C. DUPERRAY, S. HAMON, AND A. DE KOCHKO. 2005. Investigation on the origins of stoichiometric error in genome size estimation using heat experiments. Consequences on data interpretation. *Annals of Botany* 95: 111-118.

OHRI, D., R. M. FRITSCH, AND P. HANELT. 1998. Evolution of genome size in *Allium* (Alliaceae). *Plant Systematics and Evolution* 210: 57-86.

OLMO, E. 2003. Reptiles: a group of transition in the evolution of genome size and of the nucleotypic effect. *Cytogenetic and Genome Research* 101: 166-171.

ORGEL, L. E., AND F. H. C. CRICK. 1980. Selfish DNA: the ultimate parasite. *Nature* 288: 645-646.

OTTO, S. P., AND J. WHITTON. 2000. Polyploid incidence and evolution. *Annual Review of Genetics* 34: 401-437.

PAGEL, M., AND R. A. JOHNSTONE. 1992. Variation across species in the size of the nuclear genome supports the junk-DNA explanation for the C-value paradox. *Proceedings of the Royal Society of London Series B-Biological Sciences* 249: 119-124.

PEARCE, S. R. 2007. SIRE-1, a putative plant retrovirus is closely related to a legume Ty1-*copia* retrotransposon family. *Cellular & Molecular Biology Letters* 12: 120-126.

PEARCE, S. R., M. KNOX, T. H. N. ELLIS, A. J. FLAVELL, AND A. KUMAR. 2000. Pea Ty1-*copia* group retrotransposons: transpositional activity and use as markers to study genetic diversity in *Pisum*. *Molecular and General Genetics* 263: 898-907.

PEARCE, S. R., G. HARRISON, D. T. LI, J. S. HESLOPHARRISON, A. KUMAR, AND A. J. FLAVELL. 1996a. The Ty1-*copia* group retrotransposons in *Vicia* species: Copy number, sequence heterogeneity and chromosomal localisation. *Molecular & General Genetics* 250: 305-315.

PEARCE, S. R., C. STUART-ROGERS, M. R. KNOX, A. KUMAR, T. H. N. ELLIS, AND A. J. FLAVELL. 1999. Rapid isolation of plant Ty1-*copia* group retrotransposon LTR sequences for molecular marker studies. *Plant Journal* 19: 711-717.

PEARCE, S. R., U. PICH, G. HARRISON, A. J. FLAVELL, J. S. P. HESLOPHARRISON, I. SCHUBERT, AND A. KUMAR. 1996b. The Ty1-*copia* group retrotransposons of *Allium cepa* are distributed throughout the chromosomes but are enriched in the terminal heterochromatin. *Chromosome Research* 4: 357-364.

PETROV, D. A., AND D. L. HARTL. 1997. Trash DNA is what gets thrown away: high rate of DNA loss in *Drosophila*. *Gene* 205: 279-289.

PETROV, D. A., E. R. LOZOVSKAYA, AND D. L. HARTL. 1996. High intrinsic: Rate of DNA loss in *Drosophila*. *Nature* 384: 346-349.

PETROV, D. A., T. A. SANGSTER, J. S. JOHNSTON, D. L. HARTL, AND K. L. SHAW. 2000. Evidence for DNA loss as a determinant of genome size. *Science* 287: 1060-1062.

PETTERSSON, M. W. 1991. Flower herbivory and seed predation in *Silene vulgaris* (Caryophyllacae): effects of pollination and phenology. *Holarctic Ecology* 14: 45-50.

POLZ, M. F., AND C. M. CAVANAUGH. 1998. Bias in template-to-product ratios in multitemplate PCR. *Applied and Environmental Microbiology* 64: 3724-3730.

PRICE, H. J., AND K. BACHMANN. 1976. Mitotic cycle time and DNA content in annual and perennial *Microseridinae* (Compositae, Cichoriaceae). *Plant Systematics and Evolution* 126: 323-330.

PRICE, H. J., G. HODNETT, AND J. S. JOHNSTON. 2000. Sunflower (*Helianthus annuus*) leaves contain compounds that reduce nuclear propidium iodide fluorescence. *Annals of Botany* 86: 929-934.

PRICE, H. J., K. L. CHAMBERS, K. BACHMANN, AND J. RIGGS. 1983. Inheritance of nuclear 2C DNA content variation in intraspecific and interspecific hybrids of *Microseris* (Asteraceae). *American Journal of Botany* 70: 1133-1138.

PRICE, H. J., K. L. CHAMBERS, K. BACHMANN, AND J. RIGGS. 1985. Inheritance of nuclear 2C DNA content in a cross between *Microseris douglasii* and *M. bigelovii* (Asteraceae). *Biologisches Zentralblatt* 104: 269-276.

RAEYMAEKERS, L. 2000. Basic principles of quantitative PCR. *Molecular Biotechnology* 15: 115-122.

RAYBURN, A. L., AND J. A. AUGER. 1990. Genome Size Variation in *Zea-Mays Ssp Mays* Adapted to Different Altitudes. *Theoretical and Applied Genetics* 79: 470-474.

RAYBURN, A. L., D. P. BIRADAR, D. G. BULLOCK, AND L. M. MCMURPHY. 1993. Nuclear DNA content in F1 hybrids of maize. *Heredity* 70: 294-300.

RAYBURN, A. L., D. P. BIRADAR, R. L. NELSON, R. MCCLOSKEY, AND K. M. YEATER. 2004. Documenting intraspecric genome size variation in soybean. *Crop Science* 44: 261-264.

REES, H., AND R. K. J. NARAYAN. 1981. Chromosomal DNA in Higher-Plants. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 292: 569-578.

REEVES, G., D. FRANCIS, M. S. DAVIES, H. J. ROGERS, AND T. R. HODKINSON. 1998. Genome size is negatively correlated with altitude in natural populations of *Dactylis glomerata*. *Annals of Botany* 82: 99-105.

RONQUIST, F., AND J. P. HUELSENBECK. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572-1574.

SANMIGUEL, P., B. S. GAUT, A. TIKHONOV, Y. NAKIJAMA, AND J. L. BENNETZEN. 1998. The paleontology of intergene retrotransposons of maize. *Nature Genetics* 20: 43-45.

SANMIGUEL, P., A. TIKHONOV, Y. K. JIN, N. MOTCHOULSKAIA, D. ZAKHAROV, A. MELAKEBERHAN, P. S. SPRINGER, K. J. EDWARDS, M. LEE, Z. AVRAMOVA, AND J. L. BENNETZEN. 1996. Nested retrotransposons in the intergenic regions of the maize genome. *Science* 274: 765-768.

SCHMIDT, G., AND S. J. THANNHAUSER. 1945. A method for the determination of desoxyribonucleic acid, ribonucleic acid, and phosphoprotiens in animal tissues. *Journal of Biological Chemistry* 161: 83-89.

SHIRASU, K., A. H. SCHULMAN, T. LAHAYE, AND P. SCHULZE-LEFERT. 2000. A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. *Genome Research* 10: 908-915.

SILJAK-YAKOVLEV, S., M. CERBAH, J. COULAUD, V. STOIAN, S. C. BROWN, V. ZOLDOS, S. JELENIC, AND D. PAPES. 2002. Nuclear DNA content, base composition, heterochromatin and rDNA in *Picea omorika* and *Picea abies*. *Theoretical and Applied Genetics* 103: 846-854.

SISKO, M., A. IVANIC, AND B. BOHANEC. 2003. Genome size analysis in the genus *Cucurbita* and its use for the determination of interspecific hybrids obtained using the embryo-rescue technique. *Plant Science* 165: 663-669.

SMARDA, P., AND P. BURES. 2006. Intraspecific DNA content variability in *Festuca pallens* on different geographical scales and ploidy levels. *Annals of Botany* 98: 665-678.

SOKAL, R. R., AND F. J. ROHLF. 1981. Biometry - The principles and practice of statistics in biological research. W. H. Freeman and company, New York.

SOLTIS, D. E., AND P. S. SOLTIS. 1999. Polyploidy: recurrent formation and genome evolution. *Trends in Ecology & Evolution* 14: 348-352.

STEVEN, J. C., L. F. DELPH, AND E. D. BRODIE. 2007. Sexual dimorphism in the quantitative-genetic architecture of floral, leaf, and allocation traits in *Silene latifolia*. *Evolution* 61: 42-57.

STUART-ROGERS, C., AND A. J. FLAVELL. 2001. The evolution of Ty1-*copia* group retrotransposons in gymnosperms. *Molecular Biology and Evolution* 18: 155-163.

TALAVERA, S. 1990. *Silene In* S. Castroviejo, M. Lainz, and G. Lopez Gonzalez [eds.], Flora Iberica: Plantas vasculares de la peninsula Iberica e Islas Baleares, vol. 2., 313-406. Real Jardin Botanico, C.S.I.C., Madrid.

TAYLOR, D. R. 1994a. Sex-Ratio in Hybrids between *Silene alba* and *Silene dioica* - Evidence for Y-Linked Restorers. *Heredity* 73: 518-526.

TAYLOR, D. R. 1994b. The genetic-basis of sex-ratio in *Silene alba* (=*S. latifolia*). *Genetics* 136: 641-651.

TAYLOR, D. R. 1996. Parental expenditure and offspring ratios in the dioecious plant *Silene alba* (= *Silene latifolia*). *American Naturalist* 147: 870-879.

TAYLOR, D. R. 1999. Genetics of sex ratio variation among natural populations of a dioecious plant. *Evolution* 53: 55-62.

TEMSCH, E. M., AND J. GREILHUBER. 2000. Genome size variation in *Arachis hypogaea* and *A. monticola* re-evaluated. *Genome* 43: 449-451.

TEMSCH, E. M., AND J. GREILHUBER. 2001. Genome size in *Arachis duranensis*: a critical study. *Genome* 44: 826-830.

TEOH, S. B., AND H. REES. 1976. Nuclear-DNA Amounts in Populations of *Picea* and *Pinus* Species. *Heredity* 36: 123-&.

THIBAULT, J. 1998. Nuclear DNA amount in pure species and hybrid willows (*Salix*): a flow cytometric investigation. *Canadian Journal of Botany* 76: 157-165.

THOMSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN, AND D. G. HIGGINS. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 25: 4876-4882.

VAIO, M., C. MAZZELLA, V. PORRO, P. SPERANZA, B. LOPEZ-CARRO, E. ESTRAMIL, AND G. A. FOLLE. 2007. Nuclear DNA content in allopolyploid species and synthetic hybrids in the grass genus *Paspalum*. *Plant Systematics and Evolution* 265: 109-121.

VAN'T HOF, J. 1965. Relationships between mitotic cycle duration, S period duration and the average rate of DNA synthesis in the root meristem cells of several plants. *Experimental Cell Research* 39: 48-58.

VAN'T HOF, J., AND A. H. SPARROW. 1963. A relationship between DNA content, nuclear volume and minimum mitotic cycle time. *Proceedings of the National Academy of Sciences of the USA* 49: 897-902.

VEKEMANS, X., C. LEFEBVRE, J. COULAUD, S. BLAISE, W. GRUBER, S. SILJAKYAKOVLEV, AND S. C. BROWN. 1996. Variation in nuclear DNA content at the species level in *Armeria maritima*. *Hereditas* 124: 237-242.

VERNHETTES, S., M. A. GRANDBASTIEN, AND J. M. CASACUBERTA. 1998. The evolutionary analysis of the Tnt1 retrotransposon in *Nicotiana* species reveals the high variability of its regulatory sequences. *Molecular Biology and Evolution* 15: 827-836.

VICIENT, C. M., A. SUONIEMI, K. ANAMTHAMAT-JONSSON, J. TANSKANEN, A. BEHARAV, E. NEVO, AND A. H. SCHULMAN. 1999. Retrotransposon BARE-1 and its role in genome evolution in the genus *Hordeum*. *Plant Cell* 11: 1769-1784.

VINOGRADOV, A. E. 1995. Nucleotypic effect in homeotherms: Body-mass-corrected basal metabolic rate of mammals is related to genome size. *Evolution* 49: 1249-1259.

VINOGRADOV, A. E. 1997. Nucleotypic effect in homeotherms: Body-mass independent resting metabolic rate of passerine birds is related to genome size. *Evolution* 51: 220-225.

VINOGRADOV, A. E. 1998. Buffering: a possible passive-homeostasis role for redundant DNA. *Journal of Theoretical Biology* 193: 197-199.

VOS, P., R. HOGERS, M. BLEEKER, M. REIJANS, T. VANDELEE, M. HORNES, A. FRIJTERS, J. POT, J. PELEMAN, M. KUIPER, AND M. ZABEAU. 1995. Aflp - a New Technique for DNA-Fingerprinting. *Nucleic Acids Research* 23: 4407-4414.

VYSKOT, B., AND R. HOBZA. 2004. Gender in plants: sex chromosomes are emerging from the fog. *Trends in Genetics* 20: 432-438.

WAGNER, A., N. BLACKSTONE, P. CARTWRIGHT, M. DICK, B. MISOF, P. SNOW, G. P. WAGNER, J. BARTELS, M. MURTHA, AND J. PENDLETON. 1994. Surveys of Gene Families Using Polymerase Chain-Reaction - Pcr Selection and Pcr Drift. *Systematic Biology* 43: 250-261.

WANG, Z., J. L. WEBER, G. ZHONG, AND S. D. TANKSLEY. 1994. Survey of plant short tandem DNA repeats. *Theoretical and Applied Genetics* 88: 1-6.

WARTON, D. I., I. J. WRIGHT, D. S. FALSTER, AND M. WESTOBY. 2006. Bivariate line-fitting methods for allometry. *Biological Reviews* 81: 259-291.

WAUGH, R., K. MCLEAN, A. J. FLAVELL, S. R. PEARCE, A. KUMAR, B. B. T. THOMAS, AND W. POWELL. 1997. Genetic distribution of Bare-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). *Molecular & General Genetics* 253: 687-694.

WENDEL, J. F., AND S. R. WESSLER. 2000. Retrotransposon-mediated genome evolution on a local ecological scale. *Proceedings of the National Academy of Sciences of the United States of America* 97: 6250-6252.

WENDEL, J. F., R. C. CRONN, J. S. JOHNSTON, AND H. J. PRICE. 2002. Feast and famine in plant genomes. *Genetica* 115: 37-47.

WESTERGAARD, M. 1958. The mechanism of sex determination in dioecious flowering plants. *Advances in Genetics* 9: 217-281.

WHITE, S. E., L. F. HABERA, AND S. R. WESSLER. 1994. Retrotransposons in the flanking sequence of normal plant genes - a role of *copia*-like elements in the evolution of gene structure and expression. *Proceedings of the National Academy of Sciences of the USA* 91: 11792-11796.

WILLIAMS, C. G., M. M. GOODMAN, AND C. W. STUBER. 1995. Comparative Recombination Distances among *Zea-Mays L* - Inbreds, Wide Crosses and Interspecific Hybrids. *Genetics* 141: 1573-1581.

WRIGHT, J. W. 2000. Pollination, seed predation and floral character evolution in *Silene latifolia*, Rutgers University, New Brunswick.

WRIGHT, J. W., AND T. R. MEAGHER. 2004. Selection on floral characters in natural Spanish populations of *Silene latifolia*. *Journal of Evolutionary Biology* 17: 382-395.

XIONG, Y., AND T. H. EICKBUSH. 1990. Origin and Evolution of Retroelements Based Upon Their Reverse-Transcriptase Sequences. *Embo Journal* 9: 3353-3362.

YANG, Z. H. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in BioSciences* 13: 555-556.

YANG, Z. H., AND R. NIELSEN. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution* 17: 32-43.

ZIMMERMANN, K., AND J. W. MANNHALTER. 1996. Technical aspects of quantitative competitive PCR. *Biotechniques* 21: 268-&.

ZUCKERLANDL, E., AND W. HENNIG. 1995. Tracking Heterochromatin. *Chromosoma* 104: 75-83.